

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2003, 23:22:25 ; Search time 2898 Seconds

(without alignments)
10373.774 Million cell updates/sec

Title: US-10-027-450-46

Perfect score: 1033
Sequence: 1 atcattgacgacgacgtctctc.....aaaaaaaaaaaaaaaaa 1033Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.ov:*
5: gb.pat:*
6: gb.ph:*
7: gb.pl:*
8: gb.pr:*
9: gb.ro:*
10: gb.sts:*
11: gb.un:*
12: gb.vi:*
13: gb.yi:*
14: gb.yi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.yi:*
31: em.htg.hum:*
32: em.htg.in:*
33: em.htg.inv:*
34: em.htg.other:*
35: em.htg.mus:*
36: em.htg.pin:*
37: em.htg.rtd:*
38: em.htg.vtl:*
39: em.htg.vtl:*
40: em.htgo.hum:*
41: em.htgo.mus:*
42: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	462.2	44.7	131955	2 AP004676	AP004676 Oryza sat
C 2	462.2	44.7	170586	2 AP005006	AP005006 Oryza sat
C 3	460.6	44.6	339972	2 OSA307662	AJ307662 Oryza sat
C 4	210.8	20.4	787	8 AY063029	AY063029 Arabidops
C 5	210.8	20.4	970	8 AY035158	AY035158 Arabidops
C 6	210.8	20.4	45714	8 AC006224	AC006224 Arabidops
C 7	210.8	20.4	90283	8 AC004450	AC004450 Arabidops
C 8	209.2	20.3	945	8 AY087084	AY087084 Arabidops
C 9	197.8	19.1	802	8 AY117208	AY117208 Arabidops
C 10	197.8	19.1	966	8 AY063828	AY063828 Arabidops
C 11	197.8	19.1	1014	8 AY060594	AY060594 Arabidops
C 12	196.6	19.0	975	8 AY065366	AY065366 Arabidops
C 13	196.6	19.0	89517	8 AYF17J16	AYF17J16 Arabidops
C 14	159.8	15.5	10029	1 AE012835	AE012835 Chlorobiu
C 15	159.8	15.5	9830	1 AFA3BD	AFA3BD Alcaligenes
C 16	66.6	6.4	14034	1 AE005694	AE005694 Caulobact
C 17	65.2	6.3	4347	1 AB017109	AB017109 Thermus t
C 18	65.2	6.3	4347	6 E6126	E6126 L-lysine bl
C 19	64	6.2	8217	1 AF416776	AF416776 Methyloba
C 20	62.4	6.0	3057	10 MMDG2	Y14634 Mus musculu
C 21	61.2	5.9	171187	2 AC116960	AC116960 Dictyoste
C 22	59.8	5.8	1542	3 AY070834	AY070834 Drosophill
C 23	59.6	5.8	2661	8 USMLEU1A	L20832 Ustilago ma
C 24	59.6	5.7	97683	2 AC116548	AC116548 Dictyoste
C 25	58.6	5.7	42210	1 SC1C2	AL031124 Streptomy
C 26	58.6	5.7	58930	2 AC098321	AC098321 Rattus no
C 27	58.4	5.7	113193	1 AF357202	AF357202 Streptomy
C 28	57.8	5.6	13968	1 AE000901	AE000901 Methanoba
C 29	57.6	5.6	12454	1 AE002019	AE002019 Deinococc
C 30	57.6	5.6	12992	1 AE010370	AE010370 Methanopy
C 31	57.6	5.6	268147	2 AC116966	AC116966 Dictyoste
C 32	57.2	5.5	885	10 BC002080	BC002080 Mus muscu
C 33	57.2	5.5	1047	10 BC006773	BC006773 Mus muscu
C 34	57	5.5	1466	6 AR208521	AR208521 Sequence
C 35	57	5.5	1466	6 AR208522	AR208522 Sequence
C 36	57	5.5	1472	6 AR174965	AR174965 Sequence
C 37	57	5.5	1472	6 AR174966	AR174966 Sequence
C 38	57	5.5	1875	6 AR208551	AR208551 Sequence
C 39	57	5.5	1875	6 AR208552	AR208552 Sequence
C 40	57	5.5	1881	3 AF008300	AF008300 Dicotylar
C 41	57	5.5	1881	6 AR174995	AR174995 Sequence
C 42	57	5.5	1881	6 AR174996	AR174996 Sequence
C 43	56.8	5.5	125020	9 AF429315	AF429315 Homo sapl
C 44	56.8	5.5	595	6 AX090442	AX090442 Sequence
C 45	56.4	5.5	15348	1 AE007163	AE007163 Mycobacte

ALIGNMENTS

RESULT 1
AP004676/c 131955 bp DNA linear HTG 21-MAR-2002
LOCUS Oryza sativa (japonica cultivar-group) chromosome 2 clone
DEFINITION OJ1003 B06, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP004676
VERSION AP004676.1 GI:18447935
KEYWORDS HTG; HTGS-PHASE2.
SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
clone: OJ1003 B06.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1

AUTHORS	TITLE	JOURNAL	REFERENCE	COMMENT
Sasaki,T., Matsumoto,T. and Yamamoto,K.	Oryza sativa niponbare (Ga3) genomic DNA, chromosome 2, BAC clone-OJ1003.B06	Published only in Database (2002)	2 (bases 1 to 131955)	Sasaki,T., Matsumoto,T. and Yamamoto,K. Direct Submission Submitted (30-JAN-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail: tsukubakinaia.affrc.go.jp, URL: http://ryg.dna.affrc.go.jp/, Tel:81-296-38-7441, Fax:81-296-38-7468) The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data. NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.
FEATURES	source	Location/Qualifiers		
	BASE COUNT	36297 a 29034 c 28860 g 37690 t 54 others		
ORIGIN		1..131955		
		/organism="Oryza sativa (Japonica cultivar-group)"		
		/cultivar="Niponbare"		
		/db_xref="taxon:39947"		
		/chromosome="2"		
		/clone="OJ1003.B06"		
Query Match	44.7%; Score 462.2; DB 2; Length 131955;			
Best Local Similarity	78.2%; Pred. No. 5.4e-85;			
Matches 570; Conservative	1; Mismatches 149; Indels 9; Gaps 1			
27	GGCGGTGTCACGCGACGCGCTTCTAGACCCCAATCCAGCTCCACAGCGCGTTTATCCG	86		
DB 87597	GGCGGCGCGGTGACAGAGCTTCTGACACCGTGTCCACGCGCTCCAGAGCGCTTCCGCG	87538		
87	GGCGCTCCAGCAGCTACCGTGCATGCGCCCTCCACAGCTAAATAGCGCGCGCGCTGGTCCAT	146		
DB 87537	CCGCACTGGGTGGCGGCTATCTGCGCGCGCGCGCTGGAATGCGACAGCTGTCCCT	87478		
147	C-----GTCGCCGCGCGCGCTGCTGTGCGCGCGCGCGCGAGCAGCTGCGCTGTGAC	197		
DB 87477	GACCGCGGTGTCGCGCGCGCTGCGGTGCGCGTGCAGCGCGCGGAGCATCGATCGCGCG	87418		
198	CGTTTTCACGCGGAGTGTCTGTGTGGGCGGACAAATATGACACCGACCAAGATCATCC	257		
DB 87417	CGTATTCACGCGGAGTGTCTGTGTGGGAGTAATCATGACACGACCAAGATCATCC	87358		
258	CGCGGAGCAGCTACCTGTGTCCTTCACAGCGGAGAGTACCGCAAGCTCGTGGTCTT	317		
DB 87357	GGCGGAGCAGCTACCTGTGTCCTTCACAGCGGAGAGTACCGCAAGCTCGTGGTCTT	87298		
318	CGCGTTGCGGGGCTCCATTCGCGGCTACCGACGCGCTGTGCTCGCGGTAGGA	377		
DB 87297	CGCGTTGCGGGGCTCCATTCGCGGCTACCGACGCGCTGTGCTCGCGGTAGGA	87238		
378	GTCCTCCGCTAGCGCATCATGTCGCGCGGAGGCCAATTGGGGGCGGTTCCCTCGCA	437		
DB 87237	GACACCGCGCTAGCGCGCTCATCATGCGCGGAGGCCAATTGGGCTGCGGCTCTCCCGCA	87178		
438	GCAAGCGCGCGCTGCGCTTGGGCGCGCTGCGGACAGCGCATTTGTTGCSGAGGCTACGC	497		
DB 87177	GCAAGCGCGCGCTGCGCTTGGGCGCGCGCTGCGGACAGCGCATTTGTTGCSGAGGCTACGC	87118		
498	GCGCATCTTTTTCGCAACTCTGTGGGCACTGGAGAGTGTACCTCTGGAGGCTACAGGA	557		
DB 87117	GCGCATCTTTTTCGCAACTCTGTGGGCACTGGAGAGTGTACCTCTGGAGGCTACAGGA	87058		

Accession	Sequence	Length
QY 558	CGTTGGGCGCTGGAAGAGTGCACAGCGGGAGTGTGTCACCGGAGACCTGGCAATC	617
Db 87057	CACGTGACCCCTGGAAAGGAGTGCACAAACCGGGGATGTGTACGGTGAACCTTGATTAATG	86998
QY 618	CGTTTATTAATACACACACCTCTGGCCAAAGAGTACAACTGAAACCAATTTGTATGCTGG	677
Db 86997	CCTCATATCAACCAACACATCCGGCAAGCAGTACAACTGAAAGCCTATCCGGGATGCGCG	86938
QY 678	CCCTTAATTGAGCGCGGAGGATCTTTGGCTACGCCCGGAAAGAGGAATGATGCGCTC	737
Db 86937	GCCCGTTATTGAGCGAGCGCGGATCTTTGGCTATCCCGAAGACCGGATGATGCGCATC	86878
QY 738	GAAAGCTGC 746	
Db 86877	CAAGTCTGC 86869	

RESULT 2	AP005006/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
AP005006	170586 bp	DNA	linear	HTG 28-MAR-2007			
Oryza sativa (japonica cultivar-group)	chromosome 2	clone P0519E06					
*** SEQUENCING IN PROGRESS ***, in ordered pieces.							
AP005006							
AP005006.1	GI:197373546						
HTG, HTGS_PHASE2.							
Oryza sativa (japonica cultivar-group)	(cultivar:Nipponbare)	DNA,					

ORGANISM	REFERENCE
<i>Oryza sativa</i> (japonica cultivar-group)	1
Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;	
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
Ehrhartoideae; Oryzeae; <i>Oryza</i> .	

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Sasaki, T., Matsumoto, T., and Yamamoto, K.
Oryza sativa nipponbare (GAS) genomic DNA, chromosome 2, PAC
clone: F0519E06
Published Only in Database (2002)
2 (bases 1 to 170586)
Sasaki, T., Matsumoto, T., and Yamamoto, K.
Direct Submission
Submitted (27-MAR-2002) Takuji Sasaki, National Institute of
Research Program: Kannondai

COMMENT

COMMENT

NOTE: it currently consists of 1 config. Gaps between the configs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES	Location/Qualifiers
source	1. .170586 "Feature seting (temporal cultivation)"

BASE COUNT	a	c	g	t	others
ORIGIN	46926	37577	38026	47905	152

Query Match	44.7%	Score	462.2;	DB 2;	Length	170586;			
Similarity	78.2%	Pred. No.	5.7e-85;						
Matches	570;	Conservative	1;	Mismatches	149;	Indels	9;	Gaps	1

OY	27	GCGGCTTCACGGCAGCGCTTCTAGCCCAATCCGAGCTCCAACGACGCGCTTTATCCG	86
Db	30068	GCGCGGCCCGGTGACACAGATTCTTGAGCACCGCTGCCACGCCCCCTGCAGAGACTTTCGCCCG	30009

/product="putative protein"
 /protein_id="CAC39031.1"
 /db_xref="GI:14140114"
 /translation="MIVVAARGREREGAPAAARMHGKGGTTWMNRCSYTSVYVA
 VMLENGASTGSTRSRPSGSGGRMGMEPKLGGSGAGAGMGCRPSGNDGBGARASEE
 RIRSLQAGASPSPLTARYCSPACLPILHSRPRVAAAVR"
 8619. 8711
 /gene="W30ERIPDM"
 /number=1
 8712. .8792
 /gene="W30ERIPDM"
 /number=1
 8793. .9131
 /gene="W30ERIPDM"
 /number=2
 12306. .12944
 /gene="W50ERIPDM"
 12306. .12944
 /gene="W50ERIPDM"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAC39032.1"
 /db_xref="GI:14140115"
 /translation="MERGSERRRRVYAAALAVLLVFASAFVRCGDADGEDGAGA
 AEPGTGPIPLPPWKGSGSGSQSGSPDGAMRYGWMMAASPGKSGFGFYGSGRGG
 GGGGGGGGGGSRAYAFGGGYYGHPGGGPGGGGGGGGRNNGSGGIGYGVNYYGGY
 NGEPPGGGGGAGEGGGYGDDYGGGDVGNMNSKRSFNGGGGKTKQKGGNN"
 12306. .12944
 /gene="W50ERIPDM"
 /number=1
 complement(join(16037. .16301,17591. .17967))
 /gene="C75ERIPDM"
 complement(join(16037. .16301,17591. .17967))
 /gene="C75ERIPDM"
 /codon_start=1
 /product="PREG-like protein"
 /protein_id="CAC39033.1"
 /db_xref="GI:14140116"
 /translation="MAEGAEIYGESAHQDMRYYAALAGILERYAGRNDAAAPPAEIA
 AAPAPRPTAPGISVYAAARIARFAGCSPPACVYVAYIYIDRLRLRGRCLALALA
 VDSYVHRLITLAVLSAVKPMDDICYNNAVPAKYGGISLAEMNTLEVDFFGVFDLNT
 VTPPEADYCAVLQSEMCLCAAPPRRLRYCCLSDDDAGSSSSSLREAMEAS"
 complement(16037. .16301)
 /gene="C75ERIPDM"
 /number=2
 complement(16302. .17590)
 /gene="C75ERIPDM"
 /number=1
 complement(17591. .17967)
 /gene="C75ERIPDM"
 /number=1
 complement(join(20929. .21094,23560. .23733,23808. .23869,
 23964. .24200))
 /gene="C90ERIPDM"
 complement(join(20929. .21094,23560. .23733,23808. .23869,
 23964. .24200))
 /gene="C90ERIPDM"
 /codon_start=1
 /product="WRKY-like DNA-binding protein"
 /protein_id="CAC39034.1"
 /db_xref="GI:14140117"
 /translation="MPLPPPLQPCPKLLQMPFOEDOPGIHGVLSSDHCGLYPLPLP
 ALPLPSAAAATAVYALIKHSAAGSMPIGGAEEVATVTKRAGNESTGSGTMMW
 GSTMAAGRGKMKIRKKRREPRFCPTRSYVDYLDGCKRKRYGCKYKKSILHRS
 FRTHTSNCAKKVERLSTDCRMVITYTEGRHTHSPCDNSSGHEHTNCFSSP"
 complement(20929. .21094)
 /gene="C90ERIPDM"
 /number=4
 complement(21095. .23559)
 /gene="C90ERIPDM"
 /number=3
 complement(23560. .23733)
 /gene="C90ERIPDM"

```

intron      /number=3
             complement(23734, .23807)
             /gene="C90ERIPDM"
exon        /number=2
             complement(23808, .23869)
             /gene="C90ERIPDM"
             /number=2
             complement(23870, .23963)
             /gene="C90ERIPDM"
intron      /number=1
             complement(23964, .24200)
             /gene="C90ERIPDM"
exon        /number=1
             complement(join(30589, .30712,31561, .31672,31787, .31850))
             /gene="C110ERIPDM"
             complement(join(30589, .30712,31561, .31672,31787, .31850))
             /gene="C110ERIPDM"
             /codon_start=1
             /product="putative protein"
             /protein_id="CAC39035.1"
             /db_xref="gi:14140118"
             /translation="MTYEVADQNRAMKVGSAAGKGVAVNEEGQFTRRKKGIAGEL
SSPSIALRIALHAIARROGQRLTGTSKOLDPVTPLQARSAALSHWGEENVISGL"
             complement(30589, .30712)
             /gene="C110ERIPDM"
             /number=2
             complement(30713, .31560)
             /gene="C110ERIPDM"
intron      /number=2
             complement(31561, .31672)
             /gene="C110ERIPDM"
             /number=2
             complement(31672, .31787)
             /gene="C110ERIPDM"

```

[illegible]

ACCESSION AY035158.1 GI:14334969
 VERSION AY035158.1
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 970)
 REFERENCE 1
 AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishide,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shin,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Arabidopsis Full Length cDNA Clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 970)
 AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishide,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shin,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
 COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the Arabidopsis Full-length cDNA project. The cDNA was named after the collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN Arabidopsis Full-length cDNA'). Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishide,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shin,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 Yamada,K. (SSP/PGSC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGSC) contributed equally to this work as PIs.
 FEATURES
 source location/Qualifiers
 1..970
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="2"
 /clone="RAPL04-10-K01 (R09710)"
 /note="This clone is in a modified pluescript vector (lambda ZAP) as a XhoI/SstI insert."
 ecotype: Columbia"
 1..970
 /gene="At2g43090"
 1..30
 /gene="At2g43090"
 31..786
 /gene="At2g43090"
 /codon_start=1
 /evidence-experimental
 /product="putative 3-isopropylmalate dehydratase, small subunit"
 /protein_id="AAK59662.1"
 /db_xref="GI:14334970"
 /translation="MAASIGSANPTLSRLTASPKPKSEFATRSPLTFNSTVAANF KPIVSRSSSFVTRSAEPOERKTFHCIVYGDNDIDTDIITPAEFLITPSPREY EKLGSIALVGLPASTKERFVOPGEMKRTYSLITIGENFGCCSSREHAPYCLGAGACAA VVAOSYARIFERNVATGEVYPLDSEVACDECTGDVATVELREGDSLINHTTGKE YKLKPIGDAGVINDAGIIFAYARAKAMIPSA"
 CDS
 5'UTR
 gene
 3'UTR

3'UTR 787..970
 misc_feature /gene="At2g43090"
 954
 /gene="At2g43090"
 /note="artifact within poly A tail"
 BASE COUNT 269 a 210 c 211 g 280 t
 ORIGIN
 Query Match 20.4%; Score 210.8; DB 8; Length 970;
 Best Local Similarity 65.1%; Pred.No.2.2e-33;
 Matches 362; Conservative 0; Mismatches 182; Indels 12; Gaps 3;
 202 TTCACGGGAGTGTCTTGTGTCGCGACATATCGACACCGACCATTCGCCGCC 261
 ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 238 TTCATGAGACGTGTGTATGTGTGTGCGACAAACATGACACTGACCAATATCCGCCG 297
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 262 GAGCAGCTGACTGTGTGTGTCGCCGCCGACGAGTCCGACGCTGCTTCCCTGCC 321
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 298 GAGTTCTCAGCTCTGTCCTGTCGATCCAGATCGAGAACTGAGTCTTACGCT 357
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 322 TTCGGGGGGCTCCACATCCGCGCTACCCGACGCGCTTCGTCGCTCCGGTGAGAGTCC 381
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 358 TTAGTTGGCTTCCCA---GCTTCTTACAAAGAAAGATTCGTTACAGCGAGATGAAG 414
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 382 TCCCGCTACGCATCATATGTGTGCGGAGCCACACTTCGGGTGCTTCCGCGAGCAG 441
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 415 ACGAAGTACTCATCATCATATGTCGCGTGAACCTTGGATGTGATCGTACGTGAACAT 474
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 442 GCGCCGCTGCGCGCTTGGGGCGCGTGGCGACCGCATTTGCGAGGCTACGCGCGC 501
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 475 GCTCCGCTTGTGTTAGAGACAGCGGAGCTAAAGCATGTGTGCTCATGCTTATGCTAGA 534
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 502 ATCTTTTTCGCACTCCCTGCGCACCTGAGAGGTACCTTCGTGAGCTACGACGTT 561
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 535 ATCTTTTTCGAGAACCTGCTGTGCTACTGTGAGGTTATCTTTGATTT---CTGAAGTT 591
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 562 GGGGCTGAGAGAGTGCACAGACAGGGGATGTGTGTCACCGTGACCTGCTAACCCG-- 619
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 592 AGCGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 651
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 620 ----TTTTTATTAACACACACCTGTGCGACGATGATGATGATGATGATGATGATGATGAT 675
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 652 AGTATTTGATCATATATATACGACTGGGAAAGTACAACTTAAGCGGATGTGTGTGTGT 711
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 676 GGCCCTGTAATTAAGCGGAGGATCTTTCCTACGCGCGGAAAGACAGAAATGATTGGC 735
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 712 GACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 736 TCGAAGCTGCTGCAT 751
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 772 TCTGCTGCTGCTGAT 787
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 6
 AC006224/c 45714 bp DNA 1 linear PLN 27-FEB-2002
 LOCUS Arabidopsis thaliana chromosome 2 clone MFL8 map C1C10F02, complete
 DEFINITION sequence.
 ACCESSION AC006224
 VERSION AC006224.4 GI:20197626
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 45714)
 REFERENCE 1
 AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., Vanaken,S.E., Barnstead,M.E., Mason,T.M., Bowman,C.L., Rongming,C.M., Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C., Fraser,C.M. and Venter,J.C.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 45714)

AUTHORS Lin.X.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 REFERENCE 3 (bases 1 to 45714)
 AUTHORS Town,C.D. and Kaul,S.
 TITLE Direct Submission
 JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, cdcmn@tigr.org
 COMMENT On Apr 18, 2002 this sequence version replaced gi:6596536.
 FEATURES
 Source
 1..45714
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C101P02"
 /clone="MFL8"
 /complement(1..17781)
 /note="overlap with BAC clone F14B2
 (AC004450:1..53055).sequence between 11326 and 46597 in
 F14B2 was deleted in MFL8"
 /complement(433..3077)
 /gene="At2g43146"
 /note="MFL8.1"
 /complement(join<433..671,762..1004,1193..1721,1803..2021,
 2099..2166,2260..2325,2502..2583,2689..2741,2955..>3077))
 /gene="At2g43146"
 /complement(join<433..671,762..1004,1193..1721,1803..2021,
 2099..2166,2260..2325,2502..2583,2689..2741,2955..3051))
 /gene="At2g43146"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AA15157.1"
 /db_xref="GI:20197627"
 /translation="MEALSLSTFKGSLPEAIFAKKKLFFVYISGEDESDKLNRL
 ITWDASVADSLSKYCIILVHIOAGVDATNFSAIYRSPICAIAGFSGTQWREGF
 ITADIASLSEKMLGLIOETASTISFSAIAQNSQSPSSASVYLPFGSPVDA
 VASPTASVQSPSEKSTVASTSTENGDITVAKGEASPESSLCDTTKNPASVD
 GTKANVHEATEPTLRVQAKKEPIPTAPETDNTSRVSSVDRKKQGTINVEDSG
 VGVSGRDIINTKSVDTKEKPKDEGDEDEGSKKASDVHLNTRLPGSSLOKPS
 VSTLRMKYVNSNOTIGAGDLAVPRKRYTTPQDLKSLSELRFLDRALVYV
 RKRAITYQKSTYSESNNTDPNSGGFAIVKRVLSIANPSTFGGATANASSVPER
 QTRPNTVRRNNLQVGTSTFQDSEGRSNVNRKPTTSLGSIHILNHNDEAPPGDG
 NAEWNGSTQYGGSGGSDNDR"
 /complement(3457..5763)
 /gene="At2g43143"
 /note="MFL8.4: predicted by genscan and grail"
 /complement(join<3457..3658,3749..3826,3899..4276,
 4359..4674,4768..4978,5095..5763))
 /gene="At2g43143"
 /complement(join<3457..3658,3749..3826,3899..4276,
 4359..4674,4768..4978,5095..5763))
 /gene="At2g43143"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA15151.1"
 /db_xref="GI:20197632"
 /translation="MNPQOHLKLCPRKLFLEFPFLLSLYITITTKTTISSQD
 RHHPOLHAPSHSYSLPETSERNSSPPILLPPSSSSSLSYFPLCPKNTNYL
 PCHDPSTARQYSIERHRYRERHCPDIAQEKFCIVKPYGTPTFPWSPKRYAMRN
 VPFKRLAEKKTQWVRLDEGRVFPFGGSPFGVGVADVDVILSVLPASSIRVL
 DIGGVAFGAFILNYKILFMSTAPRDIHEAOVGFALBERGVLAMGLVSTYKLRPRL
 SEDVHGSRLVNMVTSYERFYPDGLIKMDVRLRLEGVYLISGPVNASRYKPKNOK
 RDSKELONOMEKLNDFVRRLCKFKIAESYVYVWKPNSHLQCRKLKALKPGICSS
 SDPDAARTKHEPECLITPLPDNDNTKTVLKNWPERLNHPKMTGSIOGTTJAGKAD
 TNLQKRVLYVDKFKFLSNKGRNVIDMAVGGFAALIKYPMVMVMVLPEDPKRN
 TLGAVVIRDLGIGTGMNCEALSTYPRYDILIHANGVFSYLDKCDIVDILLMQRIIR
 PEGAVIIRDFEDVLYVKKATITNOMRNNGTWYPRDNSVFDHGTILIVDNSIK"
 5518..5552
 /rpt_family="(TGG)n"
 /complement(6380..6652)
 /gene="At2g43136"
 repeat_region
 gene

mRNA
 CDS
 /note="MFL8.5: predicted by grail"
 /complement(<6380..>6652)
 /gene="At2g43136"
 /complement(6380..6652)
 /gene="At2g43136"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA15162.1"
 /db_xref="GI:20197633"
 /translation="MATTELKSAKYNKRRNRPSHGKRLKKKTKYITKTKSKA
 PTPDKTIEKSRSDQKTDNDEDSIIHWCTRHGAGDIRNMNNHTR"
 /complement(7042..8055)
 /gene="At2g43100"
 /note="MFL8.14: supported by cDNA: gi_16974632_gb_AY060594
 1"
 /complement(<7042..>8055)
 /gene="At2g43100"
 /complement(7264..8034)
 /gene="At2g43100"
 /codon_start=1
 /product="3-isopropylmalate dehydratase, small subunit"
 /protein_id="AA15160.1"
 /db_xref="GI:20197631"
 /translation="MAVSPTRPQALPGSSTSTSSSLATFRSPFRNGSILIPSS
 SITRGTSSPTTIPRAASESDNEALATITTHGCIYUKNDITDQITPRAACTPR
 SNOQDEDEIAHALSGLPDFHRTFTEPGENSKYSIIIGENREGSSRHAVCTG
 AAQAKIVASVARIPIFRNSVATGEVFPLESRGCECKGTGDTVIELSDSGILTN
 HTGKNYKLEKSIDAGPVIDAGGIFAYARMGMIPSLA"
 /complement(8391..9792)
 /gene="At2g43090"
 /note="MFL8.15: supported by full length cDNA:
 Ceres:31507"
 /complement(join<8391..8537,8995..>9792))
 /gene="At2g43090"
 /complement(9007..9762)
 /gene="At2g43090"
 /codon_start=1
 /product="3-isopropylmalate dehydratase, small subunit"
 /protein_id="AA15163.1"
 /db_xref="GI:20197634"
 /translation="MASSIOSANPLSLPLASPNKSSFAFRSPFLPENSVA
 KPLVSRASSPTTIRSAAPROEKTRTGCIYVGNITDQITPRAEFTIVPSNPEY
 EKIGSTALVGLPASTIKERVQPEMKTKYSIIIGENREGSSRHAVCTGAAAGAKA
 VQASVARIPIFRNSVATGEVFPLESRGCECKGTGDTVIELREGDSILINHTTGRE
 YKRLPIGDAGPVIDAGGIFAYARKGMIPSA"
 /complement(10120..12964)
 /gene="At2g43080"
 /note="MFL8.6: predicted by genscan and genefinder"
 /complement(join<10120..10227,10318..10410,10499..10588,
 10837..10888,10982..11052,11145..11219,11452..11526,
 11796..11861,12130..12182,12355..12411,12683..12730,
 12901..12964))
 /gene="At2g43080"
 /complement(join<10120..10227,10318..10410,10499..10588,
 10837..10888,10982..11052,11145..11219,11452..11526,
 11796..11861,12130..12182,12355..12411,12683..12730,
 12901..12964))
 /gene="At2g43080"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA15158.1"
 /db_xref="GI:20197628"
 /translation="MAPAMKYFGLITPVGMVIGSILQALAFIRLDESYGTGPPSL
 RGIKGRVTRLRVSWMADKDALEILGNVPEVNSPRTIYHDFLSPCECYLK
 AIAKRLQSTVVDVKTGVSVDRTSSGMFLIVENSYPILQAIKRIAVFSQVPA
 ENBELLOVLEPEQDFKPHHDVFAVTFNLKGGGVRATMLWLTLDVDEGGETYPLPA
 GDDCYCGCKIMKISYKPTKGDVAFWMSGLDGSDDPSRSHGGCEVLISGKWSATKW
 MRPKATS"
 /complement(13890..13916)
 /rpt_family="(GAA)n"
 13943..17829
 /gene="At2g43105"
 /note="MFL8.7"

mRNA	join(<13943..14092,14226..14384,14514..14625,14716..14835,14957..15111,15327..15506,15608..15642,15728..15884,16088..16141,16214..16306,16645..16787,16863..17005,17197..17217,17377..17829) /gene="At2g43105"	
CDS	join(13943..14092,14226..14384,14514..14625,14716..14835,14957..15111,15327..15506,15608..15642,15728..15884,16088..16141,16214..16306,16645..16787,16863..17005,17197..17217,17377..17486) /gene="At2g43105" /codon_start=1 /product="unknown protein" /protein_id="AAM15159.1" /db_xref="GI:20197142"	
repeat_region	translation="MSEFDPNNHRSALVILLLIGPSVAADVSWTESLESPEC TNKFMVAVLMWVDEGDFLGLAOGFALPSPDGLRPAFVDPDLSCSHLS RLDGHIALSIRGNCAFTKAKHAENASALLVINDKEDLMCKEMKEDVSLNITV LMISKSGDALNKSWDNKNEVLLIYAKRAVDLTAGLLIMAVGVVVASLWSELT DDQANESYSTIAKDVSSAGTRKDPKEKIIDISVTGAVFPIVVASIFLLIFEMS WFWVLTFFCLGSMQGMHNTIMAVILKCRHILARKSVKPLLTGMSVSLVNIICL AFAVFWIKRHTSYWVQDILGICMLTALQVNLPIKIVATVILCOAFYDIFWV ISPLIFHESVMIIVAAQDSSTGESIPMLIRIPREFDPWGDMIGFGDILPGLISF ASRVSEFTLSNPLRLILAFDLRDKIKRVTISNGYFLWLTIGIGILLTYIG LAVILGIVRGEIKELMNYGIESSESHTEDEPMPVA"	
gene	complement(18357..18435) /rpl_family="At_rpl8" complement(18779..19472) /gene="At2g43060" /note="MFL8.8" complement(<18779..>19472) /gene="At2g43060" /rpl_family="(TA)n" complement(18853..18909) complement(19002..19472) /gene="At2g43060" /codon_start=1	
mRNA	complement(<18779..>19472) /gene="At2g43060" /rpl_family="(TA)n" complement(18853..18909) complement(19002..19472) /gene="At2g43060"	
repeat_region	complement(18357..18435) /rpl_family="At_rpl8"	
gene	complement(18779..19472) /gene="At2g43060"	
mRNA	complement(<18779..>19472) /gene="At2g43060"	
repeat_region	complement(18853..18909) complement(19002..19472) /gene="At2g43060"	
CDS	complement(19002..19472) /gene="At2g43060"	
Query Match	20.4%; Score 210.8; DB 8; Length 45714;	
Best Local Similarity	65.1%; Pred. No. 5e-33;	
Matches 362; Conservative 0; Mismatches 182; Indels 12; Gaps 3;		
QY	202 TTCACGCGAGTCTTCTGTCGTCGGCGACATATGACACGACGATCATCCCGCC 261	
Db	9555 TTCATGAGACTGCTATGTCGTGGCGACATGACACGACGATTCCTCCGCG 9496	
QY	262 GAGCACTACTGTGTCCTCCACCGGACGATGACCGCAAGCTCGTCTCCGCGC 321	
Db	9495 GAGTTCTCACTCTGTCCTCCGACATGACGAGAACTCGGTTCTTACGCT 9436	
QY	322 TTCGGGGGCTCCATCCGCGGCTACCGCGCTTCGTCGCGGTCGAGAGTCC 381	
Db	9435 TTAGTGTGCTTCCA---GCTTCTTACAAGGAAGATTCGTACGCGAGATGAG 9379	
QY	382 TCCCGCTACGCGCATATGTCGCGGACCACTTCGCGTCGCTCCGCGACAC 441	
Db	9378 ACGAAGTACTCAATCATATGCGCGGTAAACCTTTGATGTGATGTACGGAAC 9319	
QY	442 GCGCGCGCGGCTTGGGGCGCGTCGGCGACGCGCCATTTGTCGAGGCTACGCGCG 501	
Db	9318 GCTCGGGTTTATTAGGAGCAGCGGAGCTAAACAGTGTGGCTCAGTCTATGCTAGA 9259	
QY	502 ATCTTTTTCGCACTCCGTCGCGCACTGAGAGGTGTACCTCTGAGACTCAGCGAGCTT 561	
Db	9258 ATCTTTTTCAGAACTCTGTGTACTGTGAGGTATATCCTTGTGAT--CTGAAGTT 9202	
QY	562 GGGCGCTGGAAGAGTGAAGACAGGAGATGTGTCACCGCGACCTTGCTAACTCG-- 619	
Db	9201 AGGGTTTGTGATGACTGACACTGGTGATGTTCGACTGTGAGTTGAGGAGGAGAT 9142	
QY	620 ----TTTTATTATACCACTCTGGCAAGAGTGAACAAGTGAACAATTTGTGATGCT 675	
Db	9141 AGTATTTGATCATCATACGACTGGGAAGAGTGAACAAGCTTAAGCCGATTTGGTATGCT 9082	
QY	676 GGCCCTGTAATGAGCGGAGGAGTCTTTCCTACGCCCGGAAGACAGAAATGATTGCG 735	
Db	9081 GGACCACTGATTGATGTGCTGATATTTGCTTATGCTAGCAAGAACTGGAATGATTCCA 9022	
QY	736 TCGAAGCTGCTGCAT 751	
Db	9021 TCTGCTGCTGCTGAT 9006	
RESULT 7	AC004450	
LOCUS	AC004450	
DEFINITION	Arabidopsis thaliana chromosome 2 clone F14B2 map CT10F02, complete sequence.	
ACCESSION	AC004450.3 GI:20197142	
VERSION	Htg.	
KEYWORDS	Arabidopsis thaliana.	
SOURCE	Arabidopsis thaliana.	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.	
REFERENCE	1 (bases 1 to 90283) Kaul, S., Shea, T. P., Fujii, C. Y., Mason, T. M., Rounsley, S. D., Lin, X., Fraser, C. M., Somerville, C. R. and Venter, J. C. Shenn, M., Ronning, C. M., Fraser, C. M., Somerville, C. R. and Venter, J. C. Unpublished	
AUTHORS	2 (bases 1 to 90283)	
JOURNAL	lin, X.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712	
JOURNAL	Medical Center Dr., Rockville, MD 20850, USA	
REFERENCE	3 (bases 1 to 90283)	
AUTHORS	Medical Center Dr., Rockville, MD 20850, USA	
JOURNAL	Town, C. D. and Kaul, S.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (27-FEB-2002) The Institute for Genomic Research, 9712	
JOURNAL	Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org	
REFERENCE	On Apr 18, 2002 this sequence version replaced gi:5598417.	
COMMENT	Location/Qualifiers	
FEATURES	1..90283	
SOURCE	/organism="Arabidopsis thaliana"	
	/cultivar="Columbia"	
	/db_xref="taxon:3702"	
	/chromosome="2"	
	/map="CT10F02"	
	/clone="F14B2"	
	complement(1..53055)	
	/note="Overlap with BAC clone MFL8 (AC006224.1..17781).sequence between 11326 and 46597 in F14B2 was deleted in MFL8"	
	complement(1..3839)	
	/gene="At2g43105"	
	/note="F14B2.1"	
	complement(join(<1..405,565..585,777..919,995..1137,1476..1568,1641..1694,1898..2054,2140..2174,2276..2455,2671..2825,2947..3066,3157..3268,3398..3556,3690..>3839))	
	/gene="At2g43105"	
	complement(join(296..405,565..585,777..919,995..1137,1476..1568,1641..1694,1898..2054,2140..2174,2276..2455,2671..2825,2947..3066,3157..3268,3398..3556,3690..>3839))	
	/gene="At2g43105"	
	/codon_start=1	
	/product="unknown protein"	
	/protein_id="AAM14939.1"	
	/db_xref="GI:20197142"	
	translation="MSEFDPNNHRSALVILLLIGPSVAADVSWTESLESPEC TNKFMVAVLMWVDEGDFLGLAOGFALPSPDGLRPAFVDPDLSCSHLS RLDGHIALSIRGNCAFTKAKHAENASALLVINDKEDLMCKEMKEDVSLNITV LMISKSGDALNKSWDNKNEVLLIYAKRAVDLTAGLLIMAVGVVVASLWSELT DDQANESYSTIAKDVSSAGTRKDPKEKIIDISVTGAVFPIVVASIFLLIFEMS WFWVLTFFCLGSMQGMHNTIMAVILKCRHILARKSVKPLLTGMSVSLVNIICL AFAVFWIKRHTSYWVQDILGICMLTALQVNLPIKIVATVILCOAFYDIFWV ISPLIFHESVMIIVAAQDSSTGESIPMLIRIPREFDPWGDMIGFGDILPGLISF ASRVSEFTLSNPLRLILAFDLRDKIKRVTISNGYFLWLTIGIGILLTYIG	

repeat_region
3866..3892
/rpt_family="(GAA)n"
4818..7662
/gene="At2g43080"
/note="F14B2.2; predicted by genscan and genefinder"
join(<4818..4881,5052..5099,5371..5427,5600..5652,
5921..5986,6256..6330,6563..6637,6730..6800,6894..6945,
7194..7283,7372..7464,7555..>7662)
/gene="At2g43080"
join(4818..4881,5052..5099,5371..5427,5600..5652,
5921..5986,6256..6330,6563..6637,6730..6800,6894..6945,
7194..7283,7372..7464,7555..7662)
/gene="At2g43080"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC64297.1"
/db_xref="GI:3763917"
/translation="MAPAKIVRGELTFVAVGAVTSGSLDIAFINLLEDSYTGFPSL
KGLRQNTIRLDVSRMANKDDELRTGVNKEPVVSWSPRITVLDFDSPECEYIK
AIAERLQVSTVVDYKGVKSDVRSQGMFLHVERSPYIIQALEKRIAVSQVPA
ENGELIQVLRPEPOQFKPHHDYFADFTNLKRGQRYATMLATLDDVSGETFEPLA
GDGCTCGKIMKGISVPTKGDVFLFWSMGLDQSDPRSIGHGCEVLSGEKSAATKW
MROKATS"
7990..9391
/gene="At2g43090"
/note="F14B2.3; supported by full length cDNA:
Ceres:31507"
join(7990..8787,9245..9391)
/gene="At2g43090"
8020..8775
/gene="At2g43090"
/codon_start=1
/product="3-isopropylmalate dehydratase, small subunit"
/protein_id="AAC64298.1"
/db_xref="GI:3763918"
/translation="MAASLOSANPTLSRTLASPKPSFATFRSPFLRENGSTSVASNF
KPLVSREASSSVTRSAEPOEKRTFGILVVDNDIDDOIIPAEILVPEPSEY
EKLGSYALVGLPASYKERFVQPCGEMKTKYSIIIGENFGCGSSREHAPVCIAGAAGA
VVAQSVARIFRNNAVGEVYPLDSEVRVCECTGDVAVELREGSITLINTTQKE
YKIRPGIDAGFVIDAGGIFAVARAKGIMPAA"
9727..10740
/gene="At2g43100"
/note="F14B2.4; supported by cDNA: gi_16974632.gb-AV060594
1"
9727..10740
/gene="At2g43100"
9748..10518
/gene="At2g43100"
/codon_start=1
/product="3-isopropylmalate dehydratase, small subunit"
/protein_id="AAC64299.1"
/db_xref="GI:3763919"
/translation="MAASLPFPQALPCSSKTKSSLATRSPFLRNGSTSLIPSSI
SITSGTSSPTIIPRAASESDSEALANTFEGICVLDNDIDDOIIPAGAACFP
SNOORDEIAHAHLSGLPDEKRTPIEPGRNRSYSIIIGENFGCGSSREHAPVGL
RAGAKAIYAESYARIFRNNAVGEVYPLDSEVRVCECTGDVITLSDSGLLTN
HTGKNYIKLSIGDAPVIDAGGIFAVARAKGIMPAA"
12144..12220
/gene="At2g43137"
12144..12220
/gene="At2g43137"
12144..12220
/product="Ath-chr2-sno27"
/note="snRNA gene
(http://rna.wustl.edu/snornaDb/Arabidopsis/)"
12301..12379
/gene="At2g43138"
12301..12379
/gene="At2g43138"
/product="Ath-chr2-sno28"
/note="snRNA gene
(http://rna.wustl.edu/snornaDb/Arabidopsis/)"
12462..12540
gene

misc_RNA
/gene="At2g43139"
12462..12540
/gene="At2g43139"
/product="Ath-chr2-sno29"
/note="snRNA gene
(http://rna.wustl.edu/snornaDb/Arabidopsis/)"
12626..12700
/gene="At2g43141"
12626..12700
/gene="At2g43141"
/product="Ath-chr2-sno28a"
/note="snRNA gene
(http://rna.wustl.edu/snornaDb/Arabidopsis/)"
14493..15102
/note="molecular marker m551"
complement(14897..14925)
/rpt_family="AT_rich"
complement(15739..15760)
/rpt_family="AT_rich"
16763..16797
/rpt_family="AT_rich"
complement(17020..17074)
/rpt_family="AT_rich"
complement(18011..18045)
/rpt_family="AT_rich"
complement(18719..19970)
/gene="At2g43130"
/note="F14B2.7; identical to GB:D01026; supported by cDNA:
gi_217838.dbj.D01026.1,ATHARA4"
complement(join(<18719..19272,19634..>19970))
/gene="At2g43130"
complement(join(18962..19272,19634..19967))
/gene="At2g43130"
/codon_start=1
/product="Ras-related GTP-binding protein (ARA-4)"
/protein_id="AAC64302.1"
/db_xref="GI:3763922"
/translation="MSDDDERGEYEFKVIILGDSAVGRSNLTRYARNEFNPNKAT
IGVFQROSMIDGKEVKAQIOTAGGERRAVTSAYRGAVGALVYDITRSFEN
VGRMLDELINHSPTVAKMLIGKKDCLESIRAVSVEGSLSEGLFPMETSAIDST
NKRAPFEMVREIYSNISRKQIANDSDSKBELTYNRSYLVKNNEGKITPCCSR"
20107..20157
/rpt_family="AT_rich"
complement(21439..24702)
/gene="At2g43140"
/note="F14B2.8; predicted by genefinder"
complement(join(<21439..21519,21599..21664,21808..21939,
22711..23012,24396..>24702))
/gene="At2g43140"
complement(join(21439..21519,21599..21664,21808..21939,
22711..23012,24396..24702))
/gene="At2g43140"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC64303.1"
/db_xref="GI:3763923"
/translation="WYPPNSKSTAHDDGADPTNOYDSACATRDPSLGFQTHHP
PPQROQHQQPNLVGHYLPGEPSSTIGFOSNASSSSSLFRHRSPPAGFDQHLPTDPN
GFSLGRPNGGGGGGDSRSLKSELFSGSSSHOEHNSLPRTSVEEAAAARNGVA
SSSMFGNNRTNMNDSNHSISFTIDQPGKRSKNSDFLETOYSKPOTTLMATMEN
LMNIPEDSVPCBARAKRGFATHPRSIARERRRPRISGKLKILQELVPMNDKQTSADM
LDLVEHKIGLOHVEYRP"
complement(26301..26225)
/rpt_family="AT_rich"
complement(29414..29451)
repeat_region
repeat_region
Query Match 20.4%; Score 210.8; DB 8; Length 90283;
Best Local Similarity 65.1%; Pred. No. 5.8e-33;
Matches 362; Conservative 0; Mismatches 182; Indels 12; Gaps 3;
QY 202 TTCACGGCGAGTGGTGGTGGTGGCGACATATCGACACCGACGACATATCCCGGC 261
DB 8227 TTCATGAGACTGTGCTATGCTGCGCGACAAATCGACACTGACCAAAATCATTCGCGC 8286

Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K., Sakurai, T., Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Full Length cDNA Clones

Unpublished

2 (bases 1 to 966)

Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission

Submitted (19-NOV-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shim, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Location/Qualifiers

1. .966

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="2"

/clone="RAFL06-82-K21 (R11883)"

/note="This clone is in a modified pBluescript vector (Lambda ZAP) as a XhoI/SstI insert.

ecotype: Columbia"

1. .966

/gene="At2g43100"

1. .14

/gene="At2g43100"

15. .785

/gene="At2g43100"

/codon_start=1

/evidence=experimental

/product="putative 3-isopropylmalate dehydratase small subunit"

/protein_id="FAL36184.1"

/db_xref="GI:17380708"

/translation="MAYSLPTFOALPCSTKTSSTIATRSPFLRNSTIIPSSI
SITSRGTSPITIPRAASESDSENEALANTTEGCLVLDNDIDTQIIPAGACFP
SNOEDREIDIAHLSGLDPFKRTFIEPGENRSKYIITIGENFGSGSREHAPVLG
AAGKAIIVASVARIFFRNSVATGVEVPLESEVCECKTGTDTVITELSDSGLLTN
HTTCKNKKTKSIGDAGVIDAGGIFAVARMGMIPSLA"

3' UTR

786. .966

/gene="At2g43100"

BASE COUNT

260 a 234 c 198 g 274 t

ORIGIN

Query Match 19.1%; Score 197.8; DB 8; Length 966;
Best Local Similarity 61.4%; Pred. No. 1.1e-30;
Matches 373; Conservative 0; Mismatches 222; Indels 12; Gaps 3;

144 CATGTCCTCCGCGCGCTCTGCTCCGCGCGGCGGACAGCTGCGCGTCAAGCCGTTT 203
182 CATCCACAGCTGCTGCGCGCTCCACAAATCCGCTTAACGAAGCCCTTAGCAACAACTT 241
204 CCACGGGAGAGCTGCTGCTGCTGCGGACATATCGACACGACAGATATCCCGCGCA 263
242 CCACGGCTCTGCTATGCTTGAAGAACAAACAGACAGCAGATATCCACGACAG 301
264 GCACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 323
302 AGCGCTTGACCTTCCATCGAACCGAACGAGCGGTGATGATGCTGCTGCTGCTGCT 361
324 CGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383
362 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
384 CCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 443
419 AAGTCACTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 478
444 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503
479 TCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
504 CTTTTCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 563
539 CTTTTCGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595
564 GCGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 619
596 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 655
620 -TTTTCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 677
656 TTTTTCGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715
678 CCCTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 737
716 ACCGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
738 GAAAGCT 744
776 ATTACT 782

RESULT 11
AY060594 1014 bp mRNA linear PLN 19-NOV-2001
LOCUS
DEFINITION
Arabidopsis thaliana At2g43100/MFL8.4 mRNA, complete cds.
ACCESSION
AY060594.1 GI:16974632
VERSION
FLI.CDNA.
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots.
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
1 (bases 1 to 1014)
Bowler, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,
Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,
Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE
JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 1014)
Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Ban, J.,
Bowler, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,
Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,
Davis, R.W., Theologis, A. and Ecker, J.R.

intron complement(749..826)
/gene="F17J16_10"
/number=1
exon complement(827..967)
/gene="F17J16_10"
/number=2
intron complement(968..1123)
/gene="F17J16_10"
/number=2
exon complement(1124..2038)
/gene="F17J16_10"
/number=3
gene 2890..4538
/gene="F17J16_20"
complement(join(2890..3456,3540..3955,4034..4336,
4433..4538))
/gene="F17J16_20"
complement(join(2890..3456,3540..3955,4034..4336,
4433..4538))
/gene="F17J16_20"
complement(join(2890..3456,3540..3955,4034..4336,
4433..4538))
/note="similarity to mitochondrial RNA splicing protein
MRS2 - Saccharomyces cerevisiae, PIR:S62064"
/codon_start=1
/product="putative protein"
/protein_id="CAB86925.1"
/db_xref="GI:7529740"
/translation="MAEGFTIGISEFEAEATGAFSTYSHHQLSDLRSSAVKDSA
QTSTSPSPPLPIHAGSAGATGAKKATGAKRMREDPTGAMVEVGCDSITIK
RASVARDRLIGPVFSSHSNLTGKNEDLLPGIARKAIVNLEVIKATVTEVLL
LDPRLPEVLPVEERLKQEPQNGNENMLQASAVOSPLDEAEAGSELPEFQVL
EIALEVCSFVDKSVAALETAMVPYLDLTKNVSTENLEYSKSLTSLRVOVK
RDELEHLDDNEMADLYLTKRNTIONQTEALAGTASNSIALPAHNTSLHLSNR
SASMTNSTNEEDVDLEMLLEAYFMQDGRKNTLIVREYIDTPEYVVICDQNR
ELIOLQITLITASEPAIAETLLASFGNITPCPLSIHGVPGRVMSVTALCTVLEW
TIGIARMKKLDSS
exon complement(2890..3456)
/gene="F17J16_20"
/number=1
intron complement(3457..3539)
/gene="F17J16_20"
/number=1
exon complement(3540..3955)
/gene="F17J16_20"
/number=2
intron complement(3956..4033)
/gene="F17J16_20"
/number=2
exon complement(4034..4336)
/gene="F17J16_20"
/number=3
intron complement(4337..4432)
/gene="F17J16_20"
/number=3
exon complement(4433..4538)
/gene="F17J16_20"
/number=4
gene 5601..10237
/gene="F17J16_30"
complement(join(5601..5954,6085..6186,6264..6481,
6560..6832,6919..7108,7227..7242,7680..8140,8590..8850,
8953..9102,9329..10237))
/gene="F17J16_30"
complement(join(5601..5954,6085..6186,6264..6481,
6560..6832,6919..7108,7227..7242,7680..8140,8590..8850,
8953..9102,9329..10237))
/note="similarity to several hypothetical proteins -
Arabidopsis thaliana
contains EGF-like domain signatures AA367-379,Prokaryotic
Membrane Lipoprotein Lipid attachment site AA369-379"
/codon_start=1
/product="putative protein"
/protein_id="CAB86926.1"

/db_xref="GI:7529741"
/translation="MDRIINLPEITICHIVSFLSAEAPASTLSKRMNLEFTVIKIL
QFDDSVKNESLKDQVDGVALPTSSVRSCSLECRREDPPHYDFNFCICALKRG
ILDLKLDICAGRRYSLEVFCTKTLVKEIGSDFGFVVDVPEDAFIPALETLILN
YIRFKDLRRCAGREKLSACVLKEIYHNMEMRMKSGNISPTLERLTISHVDLYE
CEFTIRINDPPNLTLYELSDAVPDDYPIVNLDSLVKILDLTLMVDKHYGVDDNDT
ISSNPTNLINGLRVPEMTNOSPNTQAPSPPEALPYKRIKXLTIRNDPVIGFCY
EPLPVIKRCNKLTIYIDPLPHYNEDRPKSVCHISGYSPLISCLPELQITDYSGT
PGEVQLKHFLEKLSGLKLVLSLTFRSGDKRLMLPRASSCKIKHYDSLENAL
PSLKTLLIDSVKFEYRCGCCAFQKLSACVLVESVNRNLEMDWMSGCASSQTLER
LTIDRYWAEHNLESFTFPPSLTYLDYNAHVGSYPVNLDSLVKAINLGTPTDLY
EDDDPFTSDPTNLIKGLRVNELLRLMIVVGNFKSRVOKVDLSNLPREYCHITLSF
LPTKHAALTSVLSKSNLMLKVLVNLIDDSFLLPEEELPELISRTLYELKISER
VYDMWRDRGTVLPMLKTLIDDSFTHCGTIDMFLPSFVLEIRANMEWESPDEV
SSATIKLISHTGVAAEDYPLVNNKALPEARINFTVPEKIKRYRANENMLQDDDEV
VLRFSNVYKLMNGIQANOKXLTADKLEYSVLSGCCSVPFLNNKLTIGTISEGRMOA
MPVGLLHMTDKGDCDCISREDKRSITSCPVKLEIRGRGTRKMTIKHFLV
YFPLKEKNIVYEYDPTELRFQVSKVLIIQMEEYNKLSQNVOLLITDYLSEKMTA
AGRII"
exon complement(5601..5954)
/gene="F17J16_30"
/number=1
intron complement(5955..6084)
/gene="F17J16_30"
/number=1
exon complement(6085..6186)
/gene="F17J16_30"
/number=2
intron complement(6187..6263)
/gene="F17J16_30"
/number=2
exon complement(6264..6481)
/gene="F17J16_30"
/number=3
intron complement(6482..6559)
/gene="F17J16_30"
/number=3
exon complement(6560..6832)
/gene="F17J16_30"
/number=4
intron complement(6833..6918)
/gene="F17J16_30"
/number=4
exon complement(6919..7108)
/gene="F17J16_30"
/number=5
intron complement(7109..7226)
/gene="F17J16_30"
/number=5
exon complement(7227..7242)
/gene="F17J16_30"
/number=6
intron complement(7243..7679)
/gene="F17J16_30"
/number=6
exon complement(7680..8140)
/gene="F17J16_30"
/number=7
intron complement(8141..8589)
/gene="F17J16_30"
/number=7
exon complement(8590..8850)
/gene="F17J16_30"
/number=8
intron complement(8851..8952)
/gene="F17J16_30"
/number=8
exon complement(8953..9102)
/gene="F17J16_30"

Query Match 19.0%; Score 196.6; DB 8; Length 89517;
Best Local Similarity 60.3%; Pred. No. 5e-30;
Matches 382; Conservative 1; Mismatches 235; Indels 15; Gaps 3;

QY	114	CCACTGACTAAATGCGCGCGCTGGGTGCATGTCCTCCCGCGCGCTGCTCCGCGGC	173
DB	11388	CAAAACACTTACTTCTCTCCCGCGAGATCATACAGGGGTGGCTGATCATCCCTCCCA	11329
QY	174	GGGACGACAGCTGCGCGCTGTCAGCGCTTTTTCACAGCGAGTGTGGTGGAGCA	233
DB	11328	TTCAGGCGGATCAATTAACAGAGACTTTCACAGGCTCTGCTGTCTTGAACACA	11269
QY	234	TATCGACACCGACAGATATCCCGCGAGACACTTACTGTGCTCCCTCAAGCCGA	293
DB	11268	CATCGACCGATCAAAATTAATCCCGCGAGTACGGAGCTCATCTTCGATTCAGA	11209
QY	294	CGAGTACCGAGCTGCGCTTCCCTGCGCTTGGCGGCTCCCATCCGCGCTACCCGAC	353
DB	11208	AGATGCGGAGAACTGCGCTTTCGCGCTTAACGGCTTACCA--AAATCTACACGA	11152
QY	354	GCCGTTCTGCTGCTCGGCTGAGAGTCTCCCGCTACGCAATATGTGGCGAGCCAA	413
DB	11151	AGCTTTCGTTTCCAGAGAGATGAATCAAACTCATCATCATCGCGCGCATTA	11092
QY	414	CTTCGGTGCAGTCTCTCTCCGAGACGCGCGCTTGGGCGCTGCGCGCAG	473
DB	11091	TTTCGGTTGCGATCTTCCCGGAAACGCTCAGTTTGTCTGCGCGCGCGGAGCTTA	11032
QY	474	CGCATTTGTCGAGGCTACGCGCGCATCTTTTTCGCACTCCGTCGCACTGAGGA	533
DB	11031	ACCTGTGGCGAGTGTACGCTAGATCTTTTTCAGAACTGTGTAGCTACAGTGA	10972
QY	534	GGTGTACCTCTGAGACTCAGGAGCTTGGGCGCTGAGAGATGCAAGACAGGGATG	593
DB	10971	GATTTTCCGTTGA--ATCGAGGTTAGATTTGGAGAGTGCAGAAACAGGGATGT	10915
QY	594	GGTGAC-----CGTGACCTTGCTACTCCGTTTATTAACCAACCTCTGGCA	644
DB	10914	GGGACATCGAACAAGAGAGAGGATGATTTCTATCAATCATACGAGAGGA	10855
QY	645	GGAGTACAGCTGAACCAATTTGATGTGCGCTTGAATGAGGCGGAGAGATCTT	704
DB	10854	AGATATCAACTGAACCGCTCGGATGCGGTCGCTGATCGACGCGGTGATCTT	10795
QY	705	TGCTACGCGCGAGAGAGATGATGCTC 737	
DB	10794	CGCTATGCAAGAAAGCGCGCATGATCTTC 10762	
RESULT 14			
AE012835/c			
LOCUS	AE012835	10029 bp	DNA linear BCT 01-JUL-2002
DEFINITION	Chlorobium tepidum TLS	section 56 of 194 of the complete genome.	
ACCESSION	AE012835		
VERSION	AE012835.1	GI:21646549	
KEYWORDS			
SOURCE	Chlorobium tepidum TLS		
ORGANISM	Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;		
REFERENCE	1 (bases 1 to 10029)		
AUTHORS	Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,		
	Dodson, R.J., Deboy, R., Gwinn, M.L., Nelson, W.C., Haft, D.H.,		
	Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,		
	Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D.,		
	Nierman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D.,		
	Vamathevan, J., Khouli, H., White, O., Gruber, T.M., Ketchum, K.A.,		
	Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M.		
TITLE	The complete genome sequence of Chlorobium tepidum TLS, a		
	photosynthetic, anaerobic, green-sulfur bacterium		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)		
PUBMED	12093901		
REFERENCES	2 (bases 1 to 10029)		
AUTHORS	Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M.,		
	Dodson, R.J., Deboy, R., Gwinn, M.L., Nelson, W.C., Haft, D.H.,		
	Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F.,		

TITLE	JOURNAL	REMARK
Direct Submission	Submitted (30-Apr-2002) The Institute for Genomic Research, 9712	Medical Center Dr, Rockville, MD 20850, USA
unpublished		
FEATURES		
source	1..10029	Location/Qualifiers
	/organism="Chlorobium tepidum TLS"	
	/strain="TLS"	
gene	/db_xref="taxon:194439"	
	complement(90..1694)	
	/gene="leuA-1"	
	/note="CT0612"	
CDS	complement(90..1694)	
	/gene="leuA-1"	
	/note="Identified by match to PFAM protein family HMM PF00682"	
	/codon_start=1	
	/transl_table=1	
	/product="3-isopropylmalate dehydratase"	
	/protein_id="AA071854.1"	
	/db_xref="GI:21646550"	
	/translation="WVLELYDTLRDGTGEGHINLSVODKLLIARLDEFEVDIEGG	
	WPSSNPDEEFELKARLNLKARLTAFTSTASLDNENDNLGLVRCAPVLTLE	
	GWKRAHVSFKLGTSDENAEILYRSVKELVSGRVEFDADHFDGKMDAGFAERL	
	IAAIVDGGASRVVLCIDNGSTLPHETIAALIPVRLRELIGSVGIAHANSDDLAVNSIE	
	IARVAGVOVGITINGIERGCGNLLVSLIPNLMLKIGAFESHVDLKLTSKRSVE	
	ILNLPDPSKAPFVKSAPFAKRGILHVSAAKKESSLYEHLDPMLVGNRRLVSLFAG	
	SNIRYKAOEGLSLPEKGEVFKLIVAHVKKEHOGQEPGGAASPELLRLRELQFP	
	YEVLESKVYIIONCEIKAYDAQVMKVMGDEDFQVNAQDGPVNAIDAKRAKLILH	
	IPDRMTLIDYKRVLEEKSGTSARVRLIESSDSQNSMGTVGVSTNITIASQLALN	
	DSINLYLFYMQSKAAATAPSEALNS"	
gene	/gene="CT0613"	
	complement(1703..2275)	
CDS	/gene="CT0613"	
	/note="Identified by match to PFAM protein family HMM PF00694"	
	/codon_start=1	
	/transl_table=1	
	/product="3-isopropylmalate dehydratase, small subunit,	
	putative"	
	/protein_id="AA071855.1"	
	/db_xref="GI:21646551"	
	/translation="MDRIIGKAYVYLGNKINIDDOIIPAEHLVLSLDPDEYKMGKXA	
	LSGVPIQAGLPREGNIPVYVEGEFTSPSIIITAGPFGCGSSRREHAPALVAGAKAI	
	IAESYAHIFPRNCYDGSFVYFETAPLPLKSLIMTGDLSLDMENNTLNLQNTIEL	
	RPIDGVINIVDAGSIFETARKNMLMASTEA"	
	complement(2311..3606)	
	/gene="CT0614"	
	complement(2311..3606)	
	/gene="CT0614"	
	/note="Identified by match to PFAM protein family HMM PF00330"	
	/codon_start=1	
	/transl_table=1	
	/product="3-isopropylmalate dehydratase, large subunit,	
	putative"	
	/protein_id="AA071856.1"	
	/db_xref="GI:21646552"	
	/translation="MAOTIYOKIFBARANKRFVDPGQSVMLNVDLTLHDVCGPFTD	
	IFKQEPGNALAEQSPNPGVLPVLTSHPTCSGAPMGCGSGIGNTDAFLIGTKLME	
	KVPSMKFTREGQPEPLRKDILIIDITDITDGAQYRAMPEDEAYSLPIDERMFL	
	LCNNAIEAGGNGGTLADAYTEAFVAKRSKPELFTSPDPAQIHSMRYVEMERPI	
	VAKHSPDNRAIVHSAVAGPITKSYIGSGTGLKDFPKLAALIKMKVAVATNTVPA	
	TLVAASQLETEMYDQGLRIIEFEGCNALPSCAACLGSPDITVGRSVNDVVVSTT	
	NRNPPGMSGRFASVYASPLTAAASATIGKLTDPDPL"	
	complement(3775..4833)	
gene	/gene="leuB"	

DEFINITION Alcaligenes eutrophus 4-hydroxybutyrate dehydrogenase (ybd), ORF 2 and 4-10 genes, complete cds, and ORF3 and 11, 3' end.
ACCESSION L36817
VERSION 1.0
KEYWORDS 4-hydroxybutyrate dehydrogenase; NAD-dependent; lipase.
SOURCE Alcaligenes eutrophus (strain H16 SK4040) DNA.
ORGANISM Ralstonia eutropha
Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
REFERENCE 1 (bases 1 to 9830)
AUTHORS Valentin,H.E., Zwirgmann,G., Schonebaum,A. and Steinbuechel,A.
TITLE Metabolic pathway for biosynthesis of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) from 4-hydroxybutyrate by Alcaligenes eutrophus
JOURNAL Eur. J. Biochem. 227 (1-2), 43-60 (1995)
MEDLINE 95154322
PUBMED 7851418
FEATURES
source
1. .9830 Location/Qualifiers
/organism="Ralstonia eutropha"
/strain="H16 SK4040"
/db_xref="taxon:510"
1. .164
/gene="ORF3"
/codon_start=2
/transl_table=11
/protein_id="AAC41419.1"
/db_xref="GI:703067"
/translation="NSVGNLAGFVSPYVWALKDMQTOSTQLPMYVLSAILVGAVLW LPPAKLVNR"
1. .163
/gene="ORF3"
1/0. .204
/note="putative"
222. .225
232. .1014
/note="ORF4"
/codon_start=1
/transl_table=11
/protein_id="AAC41420.1"
/db_xref="GI:695274"
/translation="MPREFANLSMNYNEHAFIDRFAAAAADGFAVEVLLPYEHAAS LRARLDTGALFNAAPDGMAGGERGLAALFGRADFRGTIGRALAEAGVIGNDRI HWAGLIPADPADRCARCTLENIAPANAAGVTLIEPIINTRDMGFEINRDD GOALCKEVGAAANKVQDCYHCQIVGDAVMKLRDIAGIGHIQLAGVDEPHEPDVGE LNPYLFYEVWDTIGYDGMICGEYRPAAGTSAGLGNLKPYLGR"
1053. .2027
/note="ORF5"
/codon_start=1
/transl_table=11
/protein_id="AAC41421.1"
/db_xref="GI:695275"
/translation="MNYLTGGAGFLGLOARLLRGTLINDGPPATIKRLTLDVY APGLDARRVVYTGDISDPAYLRQALIDTGTGAVPHLAANYSCQAEADPDIQGRVLD ASRALLETGRELGHQPRVLTSSVAYVGGULPVPVODTILANQSSYGCOKAIGELL SDYSRRGFVDRVLRPLTISVRGPKMAASSTFASGIIIRPLISGVANCAVAPETPLM LISRAVVAALVNGIELAGERLGNRRVYNLPGLSVTAAGMIELRRAAGNAVADRVTW EREARVENIVGTMPAAWNAERLALAFQSDASDEVIIRAYMEDAGLAK"
complement(2076. .2107)
/note="putative"
complement(2131. .2155)
/note="putative"
complement(2143. .2580)
/note="ORF6"
/codon_start=1
/transl_table=11
/protein_id="AAC41422.1"
/db_xref="GI:695276"
/translation="MAKEDFRHSITPLRVAEYVDQSIIVENAHYLYICDICTVEYMA LGIRYEDVLAHGVDFIVVKSSTLEVASARFDDLEIRGMSRLGSSMLFVEMFR GDEHLITGEIYVVCADPATQKSAIIPGVREIIIEGYEVVAPAG"
2590. .2595
RBS
terminator
complement(2619. .2665)

terminator
promoter
CDS
RBS
CDS
misc_signal
-35_signal
-10_signal
promoter
RBS
gene
CDS
/note="putative"
complement(2675. .2722)
/note="putative"
complement(2692. .2706)
/note="24/-12 signal"
complement(2697. .3680)
/note="ORF7"
/codon_start=1
/transl_table=11
/protein_id="AAC41423.1"
/db_xref="GI:695277"
/translation="MQRRHFIPARAGIAATAALGLAMPQAQADKFEQRPRLVYIG TAGSTDIPEPRVLADNASKILIGQVIVENKPGAGVLPAPQMOSTAPDQYTAQVAMP VYRLPYRTKINMDPVKDLNIIINAGSEGLVYPADSPILKTMQETIAYAKANQKILTY GSPGSMPTLTKIMBELAMKQGVQFSHPYKGNSESMQALLGHHVMSVADTLPAMAPVE QGKRLISTWGEKRSARFSPVPLKTEIGICITVOTSPGIVAPKGTGPKIVOKLHDAFK KAMDPNYRESLAKFDMEPYTKNSQQAAPAAETVKKRAIIEKGLAQAQ"
complement(3691. .3696)
/note="ORF8"
/codon_start=1
/transl_table=11
/product="lipase-like enzyme"
/protein_id="AAC41424.1"
/db_xref="GI:695278"
/translation="MSRCGRRAELHAAGTSGSASFPMSAPDSSPDAPATPGFSS PLDPQVALLLELILARAKRPPILHMEPEDAKIAYEKSPILIDINPPYVMAEDLLAPAR DGHAIPLIRYTPREASWTPEPLLVYFHGGGFTVGSVSDSLPCLRLCGQADCMVASV DRYGAPQMPPTANDAFDVLHFAFGRLGADPARIAGVDSAGGTLLAACAVALR NAGLAPVQLILITGTCARQDTPSHRLAGGYLLTAMIRFPQAYTLDQASDDMRF APIDGGAGAEVGTCPAMITAVAGTIDPLHDEGVAIYAEKLAAGVAAITLADYPPMIDF FKLGRFPVAVQAQAAHAEVAAALRAAFGTPHN"
4853. .4859
/note="43 region"
4864. .4869
4888. .4893
complement(4941. .4955)
/note="24/-12 signal"
5105. .5109
5120. .6268
/gene="ybd"
5120. .6268
/gene="ybd"
/EC_number="1.1.1.61"
/codon_start=1
/transl_table=11
/product="NAD-dependent 4-hydroxybutyrate dehydrogenase"
/protein_id="AAC41425.1"
/db_xref="GI:695279"
/translation="MARTYVLTHTHIDGAVSLKSECEGIRGRRPLLYTDKGVVAG VAORAIMQGLQAVNPDETSNPLEAMNVRKAAQVREAGCDGLVAVGGSSIDLARG IAILATHGELTYATTIEGSSARITDKAAPLIAYPTTSGTSGSEVAKRGAIIITLDDGRLT GHFSWHLPRKSAVCDPELTLGLPGLTAAATGMDIARCIETFLPARNPADGIALDG LERHGHTIERATRDQODDARLNMMSASQAMAFQKGLGCVHSLSHPLGGLKIDGTAT LERHGHTLNAVMPAVLRNNADAPTVYRDDRARLRLRMMHLPDGDADIAQVADMTVRGL LPTGRQMGVDEMDKVTIAGLVLDHCKTNPKBASADYRRMLEQSM"
6283. .6317
/note="putative"
6335. .6339
6345. .6866
/note="ORF2"
/codon_start=1
/transl_table=11
/protein_id="AAC41426.1"
/db_xref="GI:695280"
/translation="MSGEHLQVYVHGDITRMEYDAIVNANANGSLGGGVDAIIGAGG SAIKACRAIRIDTQGCPTGEAVITTGHLPAIPVYIAHVGVPWQGGDEGDELLANAT RNSIRLAQHHRLRLAFNPISGTIYAFPRERADIAIAAVRELAALAAPEIQVYVFCF IDENRYRVERLS"
6870. .6895
/note="putative"
complement(6870. .6895)
/note="putative"

```
terminator      6901..6927
                /note="putative"
terminator      complement(6901..6927)
                /note="putative"
CDS             complement(6908..8080)
```

Search completed: April 3, 2003, 04:00:08
Job time : 3508 secs

RBS	complement (8088 . .8092)
CDS	complement (8399 . .9025)

```

/transl_table=11
/protein_id="AAC01428.1"
/db_xref="GI:695282"
/translation="MEPTFMRGVAAPLRLINIDTDALIPSEMKRVSHHGLASGLPRL
GMRLATIDQSDPPLEFLVNLQPEYTGASILLGNSFPGGSSREHNAWALKEGIRIAYVA
PGFALIHNNCAVNCGLIPVFLPMATVQALDDCCAAATATGQVLDVLRQLEFVSPGARN
YGFLEISGROMLIEGLDPIALTLTLKASSIDAPQAGADRLRRPIHPDG"
complement(8416..8448)
terminator
CDS
complement(9027..9830)
/note="putative".

```

/translation-"EFGSSAVRSYPERVRLITCLMAVSEAGIYAPDPPIVEYLGA
 RPYAPAGAEPLALMWRSLYSDDAYVDRELSIDCQOLAIPYTWGTSQHGVAIDGCA
 VPPNAMAIDADTRQAMERALSMDIRPCQNRADDAIDAAETGSCNTRLSLDSRAAGVAG
 LAGKRAVPGTALCYPSVFAVRAAEAGLDIYVREAGFEFMRSGCSMCFVAGGESFSC
 HRQVLSSTNRNRESQSGPQTRTHLAGPATYAAASVAGRIADRRPPCA"
 complement(9032..9036)

RBS	complement(9032, .9036)			
BASE COUNT	1610 a	3352 c	3294 g	1574 t
ORIGIN				

Query Match	7.58;	Score 77.6;	DB 1;	Length 9830;
Best Local Similarity	52.9%;	Pred. No. 1.4e-05;		
Matches 164;	conservative	1;	Mismatches 145;	Indels 0;
			Gaps	0;

Oy 221 TGGTGGGCGACAAATATCGACACCAGCATCATCCCCGCCGAGCACCCTCACTTTGGTGCC 280
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8985 TGCATGCGCATCAACATCGACACCGACCATCATCCCGTCGGCGGAATGAAAGCGCGTTC 8928

Qy 281 CCTCCAGGCGGAGCAGTACCGCAAGCTGGTTCCCTTCGGCTTCGGGGGCTCCATCCG 340
||| | | | | | | | | | | | |
Db 8925 CCAAGCACGGGCTGGCCGAGGGCTGTTCGCCGGCTGGCGGCTACACTGCAGGGAGCAGACC 8866

Qy	341	CGGCGTACCGGACGCGCTGCTCGGCTCGGGGTAGAGAGTCTCCCGGCTACGCCATCATTTG	400
Db	885	GCAAGCCCGCATCCCTGTTGTTGCTGTAACCAACGCCGAGTACACCGGGCGCTGCATCTCTCC	8806

QY	401	TCGCGGAGCCACATCTCGGGTGGCGTTCCCTTCGCGAGACAGCCCGCTCGCGCTTGGGG	460
Db	8805	TGCGCGGAGCAATTTTGGCTTGGGCTCCTTCGCGGAGCATGCGGTGCGGCGCTGAAAG	8748

QY 461 CCGGTGGCGCACGGCCATGTGTGCGSAGAGGCTACGGCGGCACATCTTTTTCGAACCTCG 520
||| ||||| |||||: |||| || ||||| ||||| ||
Db 8745 AGTTTCGGCATCCGGGCCCATGTGTGCGCCCGGCTTCGGTGCATCTTCCACAACAACCTGCG 8688

QY	521	TGGCCACTGG	530
Db	8685	TGGC CACGG	8676

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2003, 22:55:49 ; Search time 284 Seconds
(without alignments)
8191.247 Million cell updates/sec

Title: US-10-027-450-46

Perfect score: 1033
Sequence: 1 atcatgagcggcgtctctc.....aaaaaaaaaaaaaaaa 1033

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001C.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032.6	100.0	1033	21	AA89462
2	961.2	93.0	995	21	AA89465
3	462.2	44.7	1112	21	AA89463
4	291.2	28.2	1107	21	AA89464
5	210.8	20.4	945	21	AA89464
6	209.2	20.3	945	21	AA89464
7	209.2	20.3	945	21	AA89464
8	196.6	19.0	829	24	ABN98576
9	65.2	6.3	4347	21	AA89462

10	63.6	6.2	217	22	AA899037	CDNA encoding for
11	57.6	5.6	297	22	AA899114	CDNA encoding for
12	57.6	5.6	387	22	AA899088	Human polynucleoti
13	57.2	5.5	403765	22	AA899683	Mycobacterium tube
14	57	5.5	1466	24	ABK82308	Diofilaria immiti
15	57	5.5	1466	24	ABK82309	Diofilaria immiti
16	57	5.5	1472	24	AA167825	D. immitis transgl
17	57	5.5	1472	24	AA167826	D. immitis transgl
18	57	5.5	1875	19	AAV33033	Diofilaria immiti
19	57	5.5	1875	19	AAV33017	Diofilaria immiti
20	57	5.5	1875	24	ABK82338	Diofilaria immiti
21	57	5.5	1875	24	ABK82339	Diofilaria immiti
22	57	5.5	1881	24	AA167855	D. immitis transgl
23	57	5.5	1881	24	AA167856	D. immitis transgl
24	56.8	5.5	595	22	AA801171	Fertilisation inde
25	56.8	5.5	823	22	ABK07664	Human ovation and
26	56.8	5.5	823	22	AA102457	Human reproductive
27	56.8	5.5	2608	21	AA102457	Human secreted pro
28	56.4	5.5	378	22	AA190863	Human polynucleoti
29	55.8	5.4	612	22	AAH71471	Human cervical can
30	55.8	5.4	404	23	ABV6394	Human prostate exp
31	55.8	5.4	435	23	ABV60921	Human prostate exp
32	55.6	5.4	464	23	ABV03810	Human prostate exp
33	55.6	5.4	465	24	AA833054	Human secreted pro
34	55.4	5.4	255	22	AAH8121	Rat differential t
35	55.4	5.4	612	22	AAH71471	Human cervical can
36	55.4	5.4	1291	21	AAV63975	CDNA from clone dd
37	55.4	5.4	1772	19	AAV63174	Human prostate exp
38	55	5.3	372	23	ABV37528	Arabidopsis thalia
39	54.8	5.3	769	21	AA828132	Human prostate exp
40	54.6	5.3	556	23	ABV40063	Human prostate exp
41	54.6	5.3	556	23	ABV40163	Human prostate exp
42	54.6	5.3	556	23	ABV42105	Human prostate exp
43	54.6	5.3	556	23	ABV43601	Human prostate exp
44	54.4	5.3	431	23	ABV04399	Human prostate exp
45	54.2	5.2	2617	24	AA170851	Human heparinase I

ALIGNMENTS

RESULT 1	AA89462	standard; CDNA: 1033 BP.
ID	AA89462	
XX	AA89462	
AC	AA89462	
XX	AA89462	
DT	15-FEB-2000	(first entry)
XX	15-FEB-2000	
DE	Corn leud subunit of 3-isopropylmalate dehydratase nucleotide sequence.	
XX	Corn leud subunit of 3-isopropylmalate dehydratase nucleotide sequence.	
XX	Corn leud subunit of 3-isopropylmalate dehydratase nucleotide sequence.	
KW	Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;	
KW	branched chain amino acid transferase; biosynthetic enzyme; antibody;	
KW	3-isopropylmalate dehydratase; ss.	
OS	Zea mays.	
XX	Zea mays.	
FH	Key	Location/Qualifiers
FT	CDS	4..753
FT		/*tag= a
FT		/product= leud
FT		/note= "Subunit of 3-isopropylmalate dehydratase"
PN	WO9921880-A2.	
XX	WO9921880-A2.	
PD	06-MAY-1999.	
XX	06-MAY-1999.	
PF	20-OCT-1998;	98WO-US22081.
XX	20-OCT-1998;	
PR	28-OCT-1997;	97US-0063423.
XX	28-OCT-1997;	
PA	(DUPO) DU PONT DE MEMOURS & CO E I.	
XX	(DUPO) DU PONT DE MEMOURS & CO E I.	

PI Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
 XX WPI: 2000-022904/02.
 DR P-PSDB; AAY28439.
 PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
 XX enzymes
 PS Claim 17; Page 94-95; 102pp; English.

CC AAX89462-X89465 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase nucleotide sequences. Sequences
 CC AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.

XX
 SX Sequence 1033 BP; 221 A; 299 C; 282 G; 230 T; 1 other:
 Query Match 100.0%; Score 1032.6; DB 21; Length 1033;
 Best Local Similarity 100.0%; Pred. No. 2.8e-209;
 Matches 1033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCATGCGCGGCTCTGTCTGGGGAGCGCGGTCTCCAGCGGCGCTTACGCCCAATC 60
 DB 1 ATCATGCGCGGCTCTGTCTGGGGAGCGCGGTCTCCAGCGGCGCTTACGCCCAATC 60
 QY 61 CGAGCTCCACACGCGGCTTATCCGGCGCTCCAGCTCACTGTCCTCCACTCA 120
 DB 61 CGAGCTCCACACGCGGCTTATCCGGCGCTCCAGCTCACTGTCCTCCACTCA 120
 OY 121 CTAAATATGCGCGCGCTGCTCATGTCCTCCGCGCGCTGTCGCGGGGGAGC 180
 DB 121 CTAAATATGCGCGCGCTGCTCATGTCCTCCGCGCGCTGTCGCGGGGGAGC 180
 OY 181 AGCTGCGCGTGTAGCGCTTTCACAGGCGAGTCTTCGAGGGGCGCAATATGAC 240
 DB 181 AGCTGCGCGTGTAGCGCTTTCACAGGCGAGTCTTCGAGGGGCGCAATATGAC 240
 OY 241 ACCGACGATCATCCCGCGCGAGCACCCTCACTGTGTCCTTCAAGCGGAGAGTAC 300
 DB 241 ACCGACGATCATCCCGCGCGAGCACCCTCACTGTGTCCTTCAAGCGGAGAGTAC 300
 OY 301 CGCAAGCTGCTCTTGTGCTTGGCGGGGCTCCCATCCGCGGCGCTACCGGCGGCTC 360
 DB 301 CGCAAGCTGCTCTTGTGCTTGGCGGGGCTCCCATCCGCGGCGCTACCGGCGGCTC 360
 OY 361 GTGCGTCCGGGAGAGAGTCTCCCGCTACGCAATCTTGTGCGGAGGAGCACTTCGGG 420
 DB 361 GTGCGTCCGGGAGAGAGTCTCCCGCTACGCAATCTTGTGCGGAGGAGCACTTCGGG 420
 OY 421 TGCGGTCTCTCGCGAGCAGCGCGCTGCGCTTGGGGCGCTGCGGAGCGGCACTT 480
 DB 421 TGCGGTCTCTCGCGAGCAGCGCGCTGCGCTTGGGGCGCTGCGGAGCGGCACTT 480
 OY 481 GTTGSAGAGGCTACGCGCGCATCTTTTTCGCAACTCCGAGGAGAGAGGTGTAC 540
 DB 481 GTTGSAGAGGCTACGCGCGCATCTTTTTCGCAACTCCGAGGAGAGAGGTGTAC 540
 OY 541 CCTGTGAGACTACGAGAGCTTGGGGCTGGAAGAGTCAAGAGAGGGAGTGTGTCAC 600
 DB 541 CCTGTGAGACTACGAGAGCTTGGGGCTGGAAGAGTCAAGAGAGGGAGTGTGTCAC 600
 OY 601 GTGAGACTTGTACTCGCTTTTATTAACACACACTCTGGCAGGAGTACAGTGA 660
 DB 601 GTGAGACTTGTACTCGCTTTTATTAACACACACTCTGGCAGGAGTACAGTGA 660

DB 601 GTGAGACTTGTACTCGCTTTTATTAACACACACTCTGGCAGGAGTACAGTGA 660
 OY 661 CCAATTGTGATGCTGCGCTGTATATTGAGCGGAGAGATCTTTGCTTACGCCGGAG 720
 DB 661 CCAATTGTGATGCTGCGCTGTATATTGAGCGGAGAGATCTTTGCTTACGCCGGAG 720
 OY 721 ACAGAGATGATGCTGCGGAGAGAGTCTGTCATGAGAGCAACTATGAGCGGCTCTG 780
 DB 721 ACAGAGATGATGCTGCGGAGAGAGTCTGTCATGAGAGCAACTATGAGCGGCTCTG 780
 OY 781 CGGAGATGAGAGATGAGTGTAGTGTAGACATTAAGATCTGTCACCTACTTATGTCGA 840
 DB 781 CGGAGATGAGAGATGAGTGTAGTGTAGACATTAAGATCTGTCACCTACTTATGTCGA 840
 OY 841 CGGAGTCTCAAAATTAAGTGTGGGCTCCGCAAAATTAAGATCAATCAATTTGGTCT 900
 DB 841 CGGAGTCTCAAAATTAAGTGTGGGCTCCGCAAAATTAAGATCAATCAATTTGGTCT 900
 OY 901 TGTACAGATGCTTTTGTGTGTACTAGTACTTGTACATTTGTACTGCTGCTACT 960
 DB 901 TGTACAGATGCTTTTGTGTGTACTAGTACTTGTACATTTGTACTGCTGCTACT 960
 OY 961 GTTCTTATCTGTTGAATTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 DB 961 GTTCTTATCTGTTGAATTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 OY 1021 AAAAAAAAAAAAAA 1033
 DB 1021 AAAAAAAAAAAAAA 1033

RESULT 2
 AAX89465
 ID AAX89465 standard; cDNA: 995 BP.
 XX AAX89465;
 AC 15-FEB-2000 (first entry)
 XX
 DE Leud subunit of 3-isopropylmalate dehydratase nucleotide sequence.
 XX
 KW Corn: soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase; ss.
 XX
 OS Trifolium aestivum.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..735
 FT /tag= a
 FT /product= Leud
 XX
 PN W0921880-A2.
 XX
 PD 06-MAY-1999.
 XX
 PF 20-OCT-1998; 98WO-US22081.
 XX
 PR 28-OCT-1997; 97US-0063423.
 XX
 PA (DUPO) DU POINT DE NEMOURS & CO E I.
 XX
 PI Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
 XX WPI: 2000-022904/02.
 DR P-PSDB; AAY28442.
 XX
 PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
 XX enzymes
 PS Claim 17; Page 98; 102pp; English.
 CC AAX89462-X89465 are fragments of corn, soybean, wheat and rice leud

QY 472 CGGCCATTGTCGSGAGGGCTACCGCGCATCTTTTTCGAACTCCGTGCACTGGA 531
DB 528 GCCGAGTGTGCGGAGTCGACGTAGATCTTTCGAACTCCGTGCACTGCGC 587
QY 532 GAGGTACCTCTGTGAGCTCAGGACGTTGGGCGCTGGAAGAGTGCACAGAGGAT 591
DB 588 GAGGTATCTCGCTAGAG--TCGGAGGAGCGCTCTGGAGAGTGCACACCGCGCAT 644
QY 592 GTGTACCGCTGAGCTTCTACTACCTCGTTTATTATTAACACACTCTGCAAGAGTAC 651
DB 645 GTGTGATCATTTGATCGGAGAGACCGCTTGATCATTCACACCGGAAAGAGAT 704
QY 652 AACCTGAACCAATTGATGTCGCTGCTGATTTGAGCGGAGGATCTTGCTTAC 711
DB 705 CGCTTGAACCGATCGGACCGGGTCCAGTGTGAGGCGCGGTGCGATCTTGCGCAT 764
QY 712 GCCGGAAGACAGATGATGCTGCTC 737
DB 765 GCCGGAAGAACCGCATGATCTCTC 790

RESULT 5
AAC47222
ID AAC47222 standard; DNA; 945 BP.
XX AAC47222;

XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 53012.
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 53012.

KM Hybridisation assay; genetic mapping; gene expression control;
KM Protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0122180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 02-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.

PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144321.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 20-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145132.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148314.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149436.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155656.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.

protein, mapping functional regions of the protein and in studying associated physiological pathways. (I) is also useful for the genetic manipulation of cells, particularly plant cells. (I) is also useful in screening assays of various plant strains to determine the strains that are best capable of withstanding a particular disease or environmental stress. (II) and (III) are useful for screening of biologically active agents, e.g. fungicides, insecticides, etc., for elucidating biochemical pathways. The screened agents are useful in improved methods of treating crops to prevent or treat disease. (II) are also useful in screening programs to identify agents that mimic or enhance the action of tolerance factors. Such agents are useful in improved methods of treating crops to enhance their tolerance to environmental stress. (I) is also useful for enhancing or inhibiting production of a biosynthetic product in a plant. (III) is useful for identifying other mediators that may induce expression of proteins of interest, for establishing the extent to which any specific insect and/or pathogen is responsible for damage to a particular plant, for identifying other mediators that enhance or induce tolerance to environmental stress, for identifying factors involved in biosynthetic pathways of nutritional, commercial, or medicinal value and for identifying productions of nutritional, commercial or medicinal value. (IV) is useful in the study of genetic function and regulation, for alteration of the cellular metabolism and for screening compounds that may affect the biological function of the gene or gene products. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docid=999909770445.

Sequence 948 BP; 245 A; 208 G; 208 G; 287 T; 0 other;

Query Match 20.3%; Score 209.2; DB 24; Length 948;

Best Local Similarity 64.9%; Pred. No. 5,7e-35; Matches 361; Conservative 0; Mismatches 183; Indels 12; Gaps 3;

202 TTCACGCGGAGTCTTGTGCTGGCGACCAATTCACACCGCAGATCATCCCGCC 261
 193 TTCACGCGGAGTCTTGTGCTGGCGACCAATTCACACCGCAGATCATCCCGCC 252
 262 GAGCAGCTACTCTGTGCTGGCGACCAATTCACACCGCAGATCATCCCGCC 321
 253 GAGTTCACGCTGCTGCTGGCGACCAATTCACACCGCAGATCATCCCGCC 312
 322 TTCGCGGGGCTCCCATCCGGCGCTACCGCGCTTCGCTCGCGGTGAGAGTCC 381
 313 TTAGTTGGCTTCCA--GCTTCTTCAAGAGACGATTCGTTACGACGAGATGAG 369
 382 TCCCGCTACGCAATTCGCTGGCGACCAATTCGCGTCTGCTCGCGAGCAC 441
 370 AGAAGTACATCATTCATTCGCGTGAACCTTGGATGATGATGATGATGATGAT 429
 442 GCGCCGCTGCGCTTGGCGCGCTGCGCGACCGCATTTGCGSAGGCGTACGCGC 501
 430 GCTCGCGTGTGTTAGGAGCAGCGCGACCTAAGCAGTGGGCTTATGCTTGA 489
 502 ATCTTTTTCGCAATTCGCTGGCGACCTGAGAGGTGATACCTGAGAGTACGAGC 561
 490 ATCTTTTTCGCAATTCGCTGGCGACCTGAGAGGTGATACCTGAGAGTACGAGC 546
 562 GCGCGCTGAGAGGATGCAACAGAGGATGCTGACCGTGGACCTTCTACTCCG-- 619
 547 AGGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 606
 620 ----TTTTTATTAACCACTTCGCAAGAGTACAGAGTGAACCAATTCGAGTGT 675
 607 AGATTTTTCATCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 666
 676 GCGCGCTGATTTGAGCGGAGGATCTTTGCTTACGCGCGGAGAGAGAGATGTCG 735
 667 GAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 736
 736 TCGAAGCTGCTGAT 751
 727 TCTGCTGCTGCTGAT 742

RESULT 8
 ABN98932
 ID ABN98932 standard; DNA: 829 BP.
 AC ABN98932;
 DT 01-AUG-2002 (first entry)
 DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 700.
 KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
 KW disease; crop; thale cress; tolerance factor; insect; pathogen;
 KW nutrition; ds.
 OS Arabidopsis thaliana.
 PN US2002023281-A1.
 PD 21-FEB-2002.
 PE 26-JAN-2001; 2001US-0770445.
 PR 27-JAN-2000; 2000US-178472P.
 PA (GORL/) GORLACH J.
 PA (ANY/) AN Y.
 PA (HAM/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (RAIN/) RAINES T M.
 PA (YUY/) YU Y.
 PA (RAME/) RAMEKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHFORD A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 PA (KRICK/) KRICKER M.
 PA (SLAT/) SLATER T.
 PA (DAVI/) DAVIS K R.
 PA (ALEX/) ALLEN K.
 PA (HOFF/) HOFFMAN N.
 PA (HURB/) HURBAN P.
 PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 PI Hurban P;
 DR WPI: 2002-400781/43.
 PT New Arabidopsis thaliana nucleic acid for identifying homologous genes,
 PT producing compositions that modulate the expression or function of its
 PT encoded protein, and mapping functional regions of protein -
 PS Claim 1: SEQ ID NO 700; 499p + Sequence Listing: English.
 CC The invention relates to an Arabidopsis thaliana nucleic acid (I)
 CC comprising a sequence capable of hybridising under stringent conditions
 CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
 CC given in the specification or its fragment. A polypeptide (II) encoded by
 CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
 CC genetically modified cell (IV) comprising an exogenous nucleic acid, is
 CC useful for screening a candidate agent for its biological effect. (I) is
 CC useful in identifying homologous or related genes, in producing
 CC compositions that modulate the expression or function of its encoded
 CC protein, mapping functional regions of the protein and in studying
 CC associated physiological pathways. (I) is also useful for the genetic
 CC manipulation of cells, particularly plant cells. (I) is also useful in
 CC screening assays of various plant strains to determine the strains that
 CC are best capable of withstanding a particular disease or environmental
 CC stress. (II) and (III) are useful for screening of biologically active

CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
 CC pathways. The screened agents are useful in improved methods of treating
 CC crops to prevent or treat disease. (ii) are also useful in screening
 CC programs to identify agents that mimic or enhance the action of tolerance
 CC factors. Such agents are useful in improved methods of treating crops to
 CC enhance their tolerance to environmental stress. (i) is also useful
 CC for enhancing or inhibiting production of a biosynthetic product in a
 CC plant. (iii) is useful for identifying other mediators that may induce
 CC expression of proteins of interest, for establishing the extent to which
 CC any specific insect and/or pathogen is responsible for damage to a
 CC particular plant, for identifying other mediators that enhance or induce
 CC tolerance to environmental stress, for identifying factors involved in
 CC biosynthetic pathways of nutritional, commercial, or medicinal value and
 CC for identifying productions of nutritional, commercial or medicinal
 CC value. (iv) is useful in the study of genetic function and regulation,
 CC for alteration of the cellular metabolism and for screening compounds
 CC that may affect the biological function of the gene or gene products.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?docid=99909770445.

XX Sequence 829 BP; 225 A; 212 C; 182 G; 210 T; 0 other;

Query Match 19.0%; Score 196.6; DB 24; Length 829;
 Best Local Similarity 60.3%; Pred. No. 2.6e-32;
 Matches 382; Conservative 1; Mismatches 235; Indels 15; Gaps 3;

114 CCACTCACTAAATGCGCGCGCTGGGCTCCATGTCGCCCGCGCGCTGCTCCGCGGC 173
 91 CAACCACTACTCTCCCTCCCGCCAGATCATACACGGGTGCTGCATCATCTCCGA 150
 174 GGGCAGACGCTCGCGCTGACGCCGTTTCCACGGGAGTCTGCTGGTGGCCACAA 233
 151 TTACGGGAGATCAATACCAAGAGACTTCCACGGCTGCTGCTCTTGAAGAACAA 210
 234 TATGACACGACGATCATCCCGCGGACCATCTGCTGGTGGCTCCACGCGGA 293
 211 CATGACACGATCAATATCCCGCGGAGTACGACATCTCATCTCCCTGATTCAGA 270
 294 CGAGTACCGCAAGCTCGGCTTCCGCTTCCGCGGGGCTCCCATCCGCGCTACCGGAC 353
 271 AGATGCGGAGAACTCGGCTTTCGGGCTTAACGGCTTACA--AAATTCACACGA 327
 354 GCGCTGCTGCTCGCGGTGAGAGTCTCCCGCTAGCCATCATCTGCGCGGAGCCAA 413
 328 ACGTTTCTGTTCCAGAGAGATCAATCAAGTACTCACTCATCTGCGCGGCACTAA 387
 414 CTGGGGTGGGCTTCCTGCGGACGCGCGCGCTTGGGCGCGGTGCGGACG 473
 388 TTTGGTGGGATCTTCCCGGACACGCTCCAGTTTGTTCGCGCGCGGAGGCTAA 447
 474 CGCCATTTGTCGAGGCTACGCGCGCATCTTTTTCGCAACTCCGTTGGCCACTGGAGA 533
 448 AGCTGTGGTGGGAGATGTAGAGATCTTTTTCGAGAACTGTAGTACAGGTGA 507
 534 GGTGACCTCTGAGCTCAGGACGTTGGGCTCGAAGAGTGCAGAGAGGAGTGT 593
 508 GATTTTCCGTTGGA---ATCGAGGTTAGATTGGACAGAGGCAAAAGAGGGATGT 564
 594 GGTGAC-----CGTGAACCTTGTACTCGTTTATTAACCAACACCTCTGGCAA 644
 565 GGTGACAAATCGAACCAAGAGAGAGGTAGATTGCTGATCAATCATACAGAGAGAA 624
 645 GAGTACAAAGCTGAACCAATTTGATGCTGACCTGTAATGAGCGGAGGAGATCTT 704
 625 AGAATACAAACTGAACCGCTCGTGTGATCGCGGTGATGACAGCGCGGTGAATCTT 684
 705 TGGCTACCGCCGGAAGACAGAGATGATGGCTG 737
 685 CGCTTATGCAAGAAAGCCGCGCATGATTCCTTC 717

RESULT 9

AAA70724
 ID AAA70724 standard; DNA; 4347 BP.
 XX
 XX AAA70724;
 AC
 XX 15-DEC-2000 (first entry)
 DT
 XX Thermus thermophilus homoconitase hydratase gene region.
 DE
 XX Genome: L-lysine biosynthesis; homoconitase hydratase; ds.
 KW
 XX Thermus thermophilus.
 OS
 XX JP2000157276-A.
 PN
 XX 13-JUN-2000.
 PD
 XX 24-NOV-1998; 98JP-0333132.
 PF
 XX 24-NOV-1998; 98JP-0333132.
 PR
 XX (AJIN) AJINOMOTO KK.
 PA
 XX WPI; 2000-485355/43.
 DR
 XX P-PSDB; AAB15394, AAB15395.
 XX
 PT L-lysine biosynthetic system gene of Thermus genus microbe -
 PS Claim 4; Page 10-12; 14pp; Japanese.
 XX

This sequence represents the genomic DNA region from Thermus thermophilus
 CC which contains the coding regions for the L-lysine biosynthetic protein
 CC homoconitase hydratase subunits I and II. The gene can be used for the
 CC generation of L-lysine-producing Thermus thermophilus strain for the
 CC production of L-lysine.

Sequence 4347 BP; 680 A; 1423 C; 1569 G; 675 T; 0 other;

Query Match 6.3%; Score 65.2; DB 21; Length 4347;
 Best Local Similarity 55.4%; Pred. No. 0.00025;
 Matches 124; Conservative 1; Mismatches 99; Indels 0; Gaps 0;

374 AGAGGCTCTCCCGTACGCAATGTTGCGGAGCAACTTGGGTCGCTCTTC 433
 1857 AGGAGGTGGGCGCCCGGAGATCTCTGTTGCGGAGCAACGGGCTCGGCTCACCC 1916
 434 GCGAGCAGCGCGCGCTTGGGCGGCTGCGGAGCGCCATCTGTTGCSAGGGCT 493
 1917 GCGAGTACGCGCCCGGAGGCTGGAAGCGCTCGGCGTCCGCAATCGCAAAAGCT 1976
 494 ACCGCGCATCTTTTTCGCACTCCGTTGGCCACGTGAGAGGTGTAACCTTGAGACTCA 553
 1977 ACCCGCGCATCTTTTTCGCACTCCGTTGGCCACGTGAGAGGTGTAACCTTGAGACTCA 553
 554 CGGAGCTTGGGCTTGAAGAGTGCAGAGAGGAGGATGTGTC 597
 2037 AGCTGTGATGCTGCTAGAGAGGAGGAGAGTGTGAGC 2080

RESULT 10
 AAS29037
 ID AAS29037 standard; cDNA; 217 BP.
 XX
 XX AAS29037;
 AC
 XX 21-NOV-2001 (first entry)
 DT
 XX cDNA encoding for human DNA-binding protein #8.
 DE
 XX Human; DNA-binding protein; histone; chromo domain protein;
 KW
 KW chromatin organisation modifier; Y-box binding protein;
 KW DNA organisation; gene transcription; malignant disease;
 KW autoimmune disorder; rheumatic disease; genetic abnormality;

KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic; ss.
XX
OS Homo sapiens.
XX WO200155162-A1.
XX
PD 02-AUG-2001.
XX
Pf 17-JAN-2001; 2001WO-US01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231966.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0238935.
PR 13-OCT-2000; 2000US-0238937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

[illegible]

DR WPI: 2001-514838/56.
P-PSDB: AAO09157.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX
PS Claim 1; SEQ ID NO 9148; 1399pp + Sequence listing; English.
XX
XX The invention relates to human polynucleotides (AA199941-AA199841) and
CC the encoded proteins (AA000010-AA01910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 387 BP; 127 A; 65 C; 63 G; 119 T; 13 other;

Query Match 5.68; Score 57.6; DB 22; Length 387;
Best Local Similarity 69.68; Pred. No. 0.0057;
Matches 78; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 922 GTTACTAGTACTGTACATTTGACTCCTGCTGCTACTGTTCTTATCTTGAATAC 981
DB 104 GTTAATAGTGTGTGAGATGGGTATTCATGTCATCTTTTGTGAAAAAAA 163

QY 982 TCGCTGTGTCACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1033
DB 164 CCTTCCTTCACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 215

RESULT 13
AA199683/c
ID AA199683 standard; DNA; 4403765 BP.
AC AA199683;
XX
XX 15-JAN-2002 (first entry)
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US6294328-B1.
PN
XX
XX 25-SEP-2001.
PD
XX 24-JUN-1998; 98US-0103840.
PR 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
PA
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
PI WPI: 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX

PS Claim 4; SEQ ID NO 2; 3pp + Sequence listing; English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 5.58; Score 57.2; DB 22; Length 4403765;
Best Local Similarity 44.88; Pred. No. 0.068;
Matches 215; Conservative 1; Mismatches 264; Indels 0; Gaps 0;

QY 18 GTCCGGGAGCGGGGAGGACGAGCGAGCGGCTTACGAGCCGAGCGAGCGAGCGG 77
DB 3927501 GCCCGGCTTGGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927442

QY 78 GTTATTCGCGCGCTCCAGCTACCTGTATTCATTCGCTTCACTAAATGCCGCCGCC 137
DB 3927441 GTTATTCGCGCGAGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927382

QY 138 TGGGTCATTCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 197
DB 3927381 CGGCGCGGCTTGAATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3927322

QY 198 CATTTCACGCGAGAGCTCTGTGTGTGGCGGACATATGACACGACCATATCC 257
DB 3927321 CGCGGTAAAGCGGAGAGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 3927262

QY 258 CGCCGAGCAGCTCACTGCTGCTGCTCCAGCGGAGCGGAGCGGAGCGGAGCGGCTTCT 317
DB 3927261 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927202

QY 318 GCGCTTCCGGGCGGCTCCATTCGCGGCGGCTTACCGAGCGGCTTGTGCTCCGGGTAGA 377
DB 3927201 CGCGGTATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927142

QY 378 GTCCCGCGTACGACATTCGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 437
DB 3927141 CGCCACCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 3927082

QY 438 GCAGCGCGCGCTCGCGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 497
DB 3927081 TGCGCGCGCTTGCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3927022

RESULT 14
ABK82308
ID ABK82308 standard; cDNA; 1466 BP.
XX
XX ABK82308;
XX
XX 27-AUG-2002 (first entry)
DE Dirofilaria immitis transglutaminase gene #5.
XX
XX
XX Nematode; transglutaminase; nematocite; Dirofilaria immitis;
KW Brugia malayi; Onchocerca volvulus; antihelmatic vaccine; gene; ss.
XX
XX Dirofilaria immitis.
OS
XX
XX US6383774-B1.
PN
XX 07-MAY-2002.
PD

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 04:00:14 ; Search time 1639 Seconds

(Without alignments)
552.846 Million cell updates/sec

Title: US-10-027-450-46

Perfect score: 1033
Sequence: 1 atcatgagcgagcgtctcttc.....aaaaaaaaaaaaaaaa 1033

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata1/pubpna/PC1_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata1/pubpna/PC1US_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032.6	100.0	1033	12	US-10-027-450-46
2	961.2	93.0	995	12	US-10-027-450-52
3	462.2	44.7	1112	12	US-10-027-450-48
4	291.2	28.2	1107	12	US-10-027-450-50
5	210	20.3	756	9	US-09-938-842A-996
6	209.2	20.3	948	10	US-09-770-445-344
7	197.8	19.1	771	9	US-09-938-842A-1335
8	196.6	19.0	829	10	US-09-770-445-700
9	187.2	18.1	369	10	US-09-878-574-157
10	96.8	9.4	373	10	US-09-878-574-1701
11	90.2	8.7	218	10	US-09-878-574-123
12	88.8	8.6	185	10	US-09-878-574-122
13	63.6	6.2	217	9	US-10-091-483-18
14	63.6	6.2	217	10	US-09-764-846-18
15	63.6	6.2	297	9	US-10-091-483-95
16	63.6	6.2	297	10	US-09-764-846-95
17	56.8	5.5	595	10	US-09-967-552A-31
18	53.6	5.2	312	10	US-09-960-352-8414
19	53.6	5.2	918	9	US-10-078-090-84

c	20	52.6	5.1	418	10	US-09-960-352-4845	Sequence 4845, Ap
c	21	52.6	5.1	458	10	US-09-770-444-478	Sequence 478, Ap
c	22	52.4	5.1	3314	10	US-09-764-864-490	Sequence 490, Ap
c	23	52.4	5.1	3320	10	US-09-764-864-31	Sequence 31, Appl
c	24	52.2	5.1	1003	9	US-10-152-661-606	Sequence 606, Appl
c	25	52.2	5.1	1003	9	US-09-866-050A-606	Sequence 606, Appl
c	26	52.2	5.1	3308	10	US-09-925-302-289	Sequence 289, Appl
c	27	51.8	5.0	1531	10	US-09-925-300-577	Sequence 577, Appl
c	28	51.6	5.0	641	9	US-09-910-664-68	Sequence 68, Appl
c	29	51.6	5.0	3320	12	US-10-047-757-1	Sequence 1, Appl
c	30	51.4	5.0	1475	9	US-10-114-893-317	Sequence 317, Appl
c	31	51.4	5.0	1605	9	US-10-001-254-13	Sequence 13, Appl
c	32	51.2	5.0	224	10	US-09-878-574-15033	Sequence 15033, A
c	33	51.2	5.0	819	9	US-09-745-763-118	Sequence 118, Appl
c	34	51.2	5.0	1308	9	US-10-151-832-1	Sequence 1, Appl
c	35	51.2	5.0	3080	9	US-10-118-984-25	Sequence 25, Appl
c	36	51.2	5.0	3080	10	US-09-728-721-25	Sequence 25, Appl
c	37	51.2	5.0	3080	12	US-10-105-931-25	Sequence 25, Appl
c	38	51	4.9	425	10	US-09-834-975-451	Sequence 451, Appl
c	39	51	4.9	1234	9	US-10-063-547-63	Sequence 63, Appl
c	40	51	4.9	1234	9	US-10-174-590-237	Sequence 237, Appl
c	41	51	4.9	1234	9	US-10-176-758-237	Sequence 237, Appl
c	42	51	4.9	1234	9	US-10-063-616-63	Sequence 63, Appl
c	43	51	4.9	1234	9	US-10-175-737-237	Sequence 237, Appl
c	44	51	4.9	1234	9	US-10-063-502-63	Sequence 63, Appl
c	45	51	4.9	1234	9	US-10-173-706-237	Sequence 237, Appl

ALIGNMENTS

RESULT 1
US-10-027-450-46
Sequence 46, Application US/10027450
Patent No. US20020102715A1
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hitz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/10/027,450
PRIOR FILING DATE: 2001-12-20
CURRENT FILING DATE: 1997 October 28
PRIOR APPLICATION NUMBER: 60/063,423
NUMBER OF SEQ ID NOS: 54
SORTNAME: Microsoft Word Version 7.0A
SEQ ID NO 46
LENGTH: 1033
TYPE: DNA
ORGANISM: Zea mays
US-10-027-450-46

Query Match 100.0%; Score 1032.6; DB 12; Length 1033;
Best Local Similarity 100.0%; Pred. No. 3.6e-275;
Matches 1033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCATGCGCGGCGCTCTCGGGAGCGGCTGTCCACGCGAGCGCTTCAGCCCAATC 60
1 ATCATGCGCGCGGCGCTCTCGGGAGCGGCTGTCCACGCGAGCGCTTCAGCCCAATC 60
DB 1 ATCATGCGCGCGGCGCTCTCGGGAGCGGCTGTCCACGCGAGCGCTTCAGCCCAATC 60
QY 61 CGAGGCTCCACGAGCGGCTTCACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 120
61 CGAGGCTCCACGAGCGGCTTCACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 120
DB 61 CGAGGCTCCACGAGCGGCTTCACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 120
QY 121 CTAAATGCGCGCGGCTTCACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 180
121 CTAAATGCGCGCGGCTTCACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 180
DB 121 CTAAATGCGCGCGGCTTCACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 180
QY 181 AACTGCGCGTGTCTGACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 240
181 AACTGCGCGTGTCTGACGCGGCTTCACGCGAGCGCTTCAGCCCAATC 240

Db 181 AGCTGCGCGTGTGACGCGCTTTTCCAGGCGAGTCTTGTGCTGGGCGCAATATTCAGC 240
QY 241 ACCGACGATATCCCGCCGAGACCTCACTGTGTCCCTCCAAAGCCGAGAGTAC 300
Db 241 ACCGACGATATCCCGCCGAGACCTCACTGTGTCCCTCCAAAGCCGAGAGTAC 300
QY 301 CGCAGCTGGTTCCTTGTGCGGAGGCTCCATCCGCGGCTACCGGAGCGCTTC 360
Db 301 CGCAGCTGGTTCCTTGTGCGGAGGCTCCATCCGCGGCTACCGGAGCGCTTC 360
QY 361 GTCCGCTCCGCTGAGAGTCTCTCCGCTACGCCATCTGTGCGGAGCGCTTCG 420
Db 361 GTCCGCTCCGCTGAGAGTCTCTCCGCTACGCCATCTGTGCGGAGCGCTTCG 420
QY 421 TCGGCTCTCTCGGAGACGACGCGCGCTGCGGCTGGGCGCGGAGCGCGCTTC 480
Db 421 TCGGCTCTCTCGGAGACGACGCGCGCTGCGGCTGGGCGCGGAGCGCGCTTC 480
QY 481 GTTGSAGAGGCTACGCGGCGATCTTTTTCGCAACTCCGCTGGCCACTGAGAGTGTAC 540
Db 481 GTTGSAGAGGCTACGCGGCGATCTTTTTCGCAACTCCGCTGGCCACTGAGAGTGTAC 540
QY 541 CCTCGGAGCTACGAGAGTGGGGCCGGAAGAGTGCAGACAGGAGTGTGTAC 600
Db 541 CCTCGGAGCTACGAGAGTGGGGCCGGAAGAGTGCAGACAGGAGTGTGTAC 600
QY 601 GTGACCTTGTCTAATCTCGTTTATTAAACCACTCTGSCAAGAGTACAAAGCTGAAA 660
Db 601 GTGACCTTGTCTAATCTCGTTTATTAAACCACTCTGSCAAGAGTACAAAGCTGAAA 660
QY 661 CCAATTTGGTATGCTGGCCCTGTAAATTGAGGCGGAGGAGTCTTGGCTACGCCCGGAAG 720
Db 661 CCAATTTGGTATGCTGGCCCTGTAAATTGAGGCGGAGGAGTCTTGGCTACGCCCGGAAG 720
QY 721 ACAGGAATGATTCGCTGGAAGCTGTGCATGAGGAGAAAGCTTATGACGCGGCTCTG 780
Db 721 ACAGGAATGATTCGCTGGAAGCTGTGCATGAGGAGAAAGCTTATGACGCGGCTCTG 780
QY 781 CGGAGATGAGAGTACCTGAGTGTAGAGTATGAGTCTGACCTACTTATGTCGA 840
Db 781 CGGAGATGAGAGTACCTGAGTGTAGAGTATGAGTCTGACCTACTTATGTCGA 840
QY 841 CGGTGTCTCAAAATAGTTGGGCTTACCGAAATTTATGATCAATCAATTTGCTT 900
Db 841 CGGTGTCTCAAAATAGTTGGGCTTACCGAAATTTATGATCAATCAATTTGCTT 900
QY 901 TGTCAAGATGCTTTTGTGTACTAGTACTGTCAATGTACTCCGCTGCTACT 960
Db 901 TGTCAAGATGCTTTTGTGTACTAGTACTGTCAATGTACTCCGCTGCTACT 960
QY 961 GTTCTATCTGTTGAAATAGTCTGTGTGCGCAAAAAAAAAAAAAAAAAAAAA 1020
Db 961 GTTCTATCTGTTGAAATAGTCTGTGTGCGCAAAAAAAAAAAAAAAAAAAAA 1020
QY 1021 AAAAAAAAAAAAA 1033
Db 1021 AAAAAAAAAAAAA 1033

RESULT 2
US-10-027-450-52
; Sequence 52, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
; APPLICANT: Faico, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/10/027,450
; CURRENT FILING DATE: 2001-12-20

;; PRIOR APPLICATION NUMBER: 60/063,423
;; NUMBER OF SEQ ID NOS: 34
;; PUBLICATION DATE: 1997 October 28
;; SOFTWARE: Microsoft Word Version 7.0A
;; SEQ ID NO 52
;; LENGTH: 995
;; TYPE: DNA
;; ORGANISM: Trifolium aestivum
US-10-027-450-52

Query Match 93.0%; Score 961.2; DB 12; Length 995;
Best Local Similarity 99.0%; Pred. No. 1,8e-255;
Matches 978; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

QY 26 CGGGGATGTCACGCGGAGCGCTTCTAGCCCATCCGAGCTCCAAACGAGCGGTTATCC 85
Db 8 CGGGGATGTCACGCGGAGCGCTTCTAGCCCATCCGAGCTCCAAACGAGCGGTTATCC 67
QY 86 GGGCTCCCACTACCTGTATCGCTCCCTCACTCACTAAATGCGCGCGCTGGTCCA 145
Db 68 GGGCTCCCACTACCTGTATCGCTCCCTCACTCACTAAATGCGCGCGCTGGTCCA 127
QY 146 TCGTCCCGCGCGCGCTGCTGCGCGCGCGCGCGAGCTGCGCTGCGCGCTGCTTCC 205
Db 128 TCGTCCCGCGCGCGCTGCTGCGCGCGCGCGAGCTGCGCTGCGCGCTGCTTCC 187
QY 206 ACGGCGAGTCTTGT 265
Db 188 ACGGCGAGTCTTGT 247
QY 286 ACGTACCTCTGT 325
Db 248 ACGTACCTCTGT 307
QY 326 CGGGGCTCCCACTGCGCGCGCTTACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 385
Db 308 CGGGGCTCCCACTGCGCGCGCTTACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 367
QY 386 GCTGCGGATGATGT 445
Db 368 GCTGCGGATGATGT 427
QY 446 CCGTGTGCTTGGGCGCTGTGCGCGAGCGGCTTGTGTGTGTGTGTGTGTGTGTGTGT 505
Db 428 CCGTGTGCTTGGGCGCTGTGCGCGAGCGGCTTGTGTGTGTGTGTGTGTGTGTGTGT 487
QY 506 TTTTTCGCAATCGT 565
Db 488 TTTTTCGCAATCGT 547
QY 566 CCGGAGAGTGTGCAAGAGAGGATGTGTCAACCGTGTGTGTGTGTGTGTGTGTGTGT 625
Db 548 CCGGAGAGTGTGCAAGAGAGGATGTGTCAACCGTGTGTGTGTGTGTGTGTGTGTGT 607
QY 626 TTAACCACTCTGTGGAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 685
Db 608 TTAACCACTCTGTGGAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 667
QY 686 TTTGAGCGCGGAGGATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 745
Db 668 TTTGAGCGCGGAGGATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 727
QY 746 CTGCATGAGGAA-----ACCTATGTGAGCGCGCTGCGGAGTGTGAAAGTAAAGT 800
Db 728 CTGCATGAGGAAAGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 787
QY 801 GGAGTTAGAGTAAAGTACTGTGACCTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 860
Db 788 GGAGTTAGAGTAAAGTACTGTGACCTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 847
QY 861 CGGCTTACCGAAATTTATGATGATCAATCAATTTGTGTGTGTGTGTGTGTGTGTGTGT 920
Db 848 CGGCTTACCGAAATTTATGATGATCAATCAATTTGTGTGTGTGTGTGTGTGTGTGTGT 907

OY	921	TGTTACTATGACTTTTCAAAATGACACCGCCGCTGACTGTTCTATCTGTTGAATA	980
Db	908	TGTTACTATGACTTTTCAAAATGACACCGCCGCTGACTGTTCTATCTGTTGAATA	967
OY	981	CTGCTCTGTGTCGCAAAAAAAAAAAAA	1008
Db	968	CTGCTCTGTGTCGCAATCAAAAAAAAA	995

Query	Subject	Score	DB	Length	1112
US-10-027-450-48	Sequence 48, Application US/10027450	44.7%	DB 12	Length 1112	
US-10-027-450-48	Patent No. US20020102715A1	78.2%	Pred. No. 1.3e-117		
US-10-027-450-48	GENERAL INFORMATION:	1	Mismatches 149	Indels 9	Gaps 1
US-10-027-450-48	APPLICANT: Falco, Saverio Carl				
US-10-027-450-48	APPLICANT: Hitz, William D.				
US-10-027-450-48	APPLICANT: Kinney, Anthony J.				
US-10-027-450-48	APPLICANT: Kahoou, Rebecca E.				
US-10-027-450-48	APPLICANT: Rafalski, J. Antoni				
US-10-027-450-48	TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES				
US-10-027-450-48	FILE REFERENCE: BB-1126				
US-10-027-450-48	CURRENT APPLICATION NUMBER: US/10/027,450				
US-10-027-450-48	CURRENT FILING DATE: 2001-12-20				
US-10-027-450-48	PRIOR APPLICATION NUMBER: 60/065,423				
US-10-027-450-48	PRIOR FILING DATE: 1997 October 28				
US-10-027-450-48	NUMBER OF SEQ ID NOS: 54				
US-10-027-450-48	SOFTWARE: Microsoft Word Version 7.0A				
US-10-027-450-48	SEQ ID NO 48				
US-10-027-450-48	LENGTH: 1112				
US-10-027-450-48	TYPE: DNA				
US-10-027-450-48	ORGANISM: Oryza sativa				
Query Match	44.7%	Score 462.2	DB 12	Length 1112	
Best Local Similarity	78.2%	Pred. No. 1.3e-117			
Matches 570	Conservative 1	Mismatches 149	Indels 9	Gaps 1	
27	GCGGATGTCACGCGACGCTTCTAGCCCAATCCGAGCTCCACACAGCGCGTTTATCCG 86				
72	GCGGCGCGCGGCGGAGACAGTTCTGGACCGCGTCCGACCGCCCTCGAGAGAGTTCCGCGC 131				
87	GCGCTCCAGACTACTGTGATCGCTCCACTCACTAAATGCCGCGCGCTGGTTCAT 146				
132	CCGACGTGGGTGCGCGGTATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 191				
147	C-----GTCCCG 197				
192	GACCGCGGTGTGCG 251				
198	CGTTTTCACAGGGAATGCTGT 257				
252	CGATTTCACAGGGAATGCTGT 311				
258	CGCGGACACCTCACTCTGT 317				
312	GCGCGGACACCTCACTCTGT 371				
318	CGCGTTTCACAGGGAATGCTGT 377				
372	CGCGTTTCACAGGGAATGCTGT 431				
378	GTCTTCGCGTACGCGCATATGT 437				
432	GACCGCGCGGTACGCGCATATGT 491				
438	GACCGCGCGGTACGCGCATATGT 497				
492	GACCGCGCGGTACGCGCATATGT 551				
498	GCGCATCTTTTTCGCAATCTCGT 557				
552	GCGCATCTTTTTCGCAATCTCGT 611				

QY	558	CGTGGGGGCGCTGGAGAGAGTCAAGACAGGGGATGTGTGTCACCGTGGACCTTGGTACTC	617
Db	612	CACCTGAGGCTGTGAAGAGAGTCCAAAGACCAGGGGGATGTGTCTACGCTGGAACTTGATTAATTG	671
QY	618	CGTTTTTATTAACCAACACCTCTGGCAAGAGGTACAAGCTGAACACCAATTTGGTGTGCTGG	677
Db	672	CGTCATGATGATCAACACACACATCCGGCAGAGCAAGTACAAAGCTTGAAAGGCTATCCGGCGATGGCGG	731
QY	678	CCCTGTAATTGAGCGGCGGAGGGATCTTTTGCCTACGCCCGGGAAGACAGGAATGATTGCGTC	737
Db	732	GCCGGTATTGAGGCGACGCGGGATCTTTGCCCTATGCCCGGGAAGACAGGAATGATCGCATC	791
QY	738	GAAAGCTGC	746
Db	792	CAAGTCTGC	800

```

RESULT 4
US-10-027-450-50
; Sequence 50. Application US/10027450
; Patent NO. US20020102715A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hltz, William D.
; APPLICANT: Kliney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/10/027,450
; PRIORITY FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 60/065,423
; PRIOR FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 50
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Glycine max
; US-10-027-450-50

```

```

RESULT 5
US-09-938-842A-996
; Sequence 996, Application US//09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME. AND METHODS OF USE
; FILE REFERENCE: S001P300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227, 866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264, 647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300, 111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 996
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-996

```

OY	562	GGGGCTGGAAAGAGTCAAGACAGGGGATATGTGTACCGGAGACCTTGTAACTCG--	619
Db	562	AGGGTTTGTGATGAGGTGACAACTGGTGATGTTCCGACTTTGAGTTGAGGGAAGGAT	621
OY	620	----TTTATTAATCAACACCTCTGGCAAGAGTACAAGCTGAACAATTTGTGATGCT	675
Db	622	AGTATTTGATTCATTCATACAGCTGGGAAAGATACAAGCTTAAGCCGATTTGGGATGCT	681
OY	676	GGCCCTGTAATTGAGCGCGGAGGAGATCTTCCCTACGCCGCGAAGACAGGAATGATTCG	735
Db	682	GGACCAAGGATTAAGCTGGTGGTATATTTCCTTAAGCTAAGAAAGCTCGAATATTCCA	741
OY	736	TGCAAGCTGCT	747
Db	742	TCTGCTGCTGCT	753

QY	DB	Matches	Best Local Similarity	Conservative	Score	DB 10	Length	Indels	Gaps
QY	202	TTCCACGCGAGCTTCCTGCTGGTGCGGACATATATGACACCGACAGATATCATCCCGCC	20.3%	20.3%	209.2	DB 10	948		
DB	193	TTCCATGAGACTGTGCTGTGTCGTGGCGACACATGACACTACCAATATTCCTCCGCG	64.9%	0	183		12		3
QY	262	GAGACACGCACTGTGCTGCTCCACGCGGAGTACCGAGTACCGAGTCTGGTCTTCTTGCC							
DB	253	GAGTTTCTCACTCTCTGTCCTTCGATTCACAGAGAAATACAGAAACCTGTTTACGCT							
QY	322	TTCCGCGGAGGCTCCCATCCGCGGCGCTTACCGACGCGCTTCCTGCTCGCTCGGCTGAGGATCC							
DB	313	TTAGTGTGGTCTTCCA--GCTTTTACAAAGAACGATTCGTTGACGCGAGGTGAGATGAAG							
QY	382	TTCCGCGTACGCGCATATTTGTGCGGCGAGGACCAACTTGGGTTGCGGTTCTCTCGCGAGACAC							

Db 370 ACAGAACTACTCATCATCATTTGGCCGTGAACAACTTGGATGTGATCTCACGTGAACAT 429
QY 442 GCGCCCGTGGCGCTTGGGGCCGCGCACCGCCATTTGGSGAGGCTACGCCGC 501
Db 430 GCTCCCGTTGTTAGGACACAGCGGAGCTAAAGCAGTGTGGCTCATCTTTATGATGA 489
QY 502 ATCTTTTTCGCACTCCGTGGCCACTGTGAGAGTGTACCCCTCGAGAGCTCAGGAGAT 561
Db 490 ATCTTTTTCAGGAACCTCTTCTACTGTGTAGGTTTATCTTTGATTT---CTGAAGTT 546
QY 562 GGGCCCTGGAAGAGTGAAGACAGCGGAGTGTGTACCGCTGAGACCTTGTCTAACTCCG-- 619
Db 547 AGGGTTTGTGATGATGTGCAACCTGTGTGTGGCTGTGAGCTGTGAGTTGAGGGAAGAGAT 606
QY 620 ----TTTTTATTACCACTCTGTGCAAGAGTGTACAACTGAACCAATTGGTGTATGCT 675
Db 607 AGTATTTTGTATCATCATCATGACACTGGGAAGAGTATTAAGCTTAAGCCGATTGTGATGCT 666
QY 676 GGGCCCTGTATTTAGGCGGAGGAGATCTTTGCTTACGCCCGGAGACAGCAATGATGCG 735
Db 667 GGAACAGTATGTATGCTGTGTATATTGCTTATGCTAGGAAAGCTGGAATGATTCCA 726
QY 736 TCGAAGCTGTGCTGAT 751
Db 727 TCTGTGCTGTGCTTGAAT 742

RESULT 7

US-09-938-842A-1335
; Sequence 1335, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: S001300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1335
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1335

Query Match 19.1%; Score 197.8; DB 9; Length 771;
Best Local Similarity 61.4%; Pred. No. 1,1e-44;
Matches 373; Conservative 0; Mismatches 222; Indels 12; Gaps 3;

QY 144 CATGTCCCGCGCGCGCTGTCCGCGCGGCGGAGCACTGCGCGCTGAGCGGTTT 203
Db 168 CATCCACGTCGTGCGCGCTCAGAAATCCGACTCTAAGCAAGCCCTAGCCACACAACCTT 227
QY 204 CCAAGGAGAGTGTGCTGTGCGGAGGAGCAATTCAGACCGCAAGCATATCCCGCGGA 263
Db 228 CCAAGGCTGTGCTGTATGCTTGAAGACAACTAGACACCGACGATCATCCAGAGG 287
QY 264 GCACTCACTCTGTGCTCCCTCCAAAGCGGAGAGTACCGCAAGCTGCTTCTTCGCTT 323
Db 288 AGCGGCTTGCACCTTCCATTCAGAACAGAGCGGTATGATCCGCTCACGCTCT 347
QY 324 CGGGGGCTCCATCCGCGGCGCTTACCCGCGCGCTTGCCTCCGGGTGAGGAGTCTC 383

Db 348 CTCTGTCTACCA---GACTTCACAAAACACGCTTCAATTGAGCCAGAGACAGATC 404
QY 384 CCGTACGCCATCATTTGTGGCGGAGGCCAACCTTGGGTGCGTTCTCTCGGAGCAGC 443
Db 405 AAGTACTCATCATTAATTCAGGCGCGGAGAACTTTGGTGGGATGTCACGTGAACATGC 464
QY 444 GCGCGCTGGCGCTGGGCGCGCTGGCGGACCGCATTTGTCGAGGAGGCTACGCGGCGAT 503
Db 465 TCCGGTCTGTCTTGGAGACAGCTGGAGCTAAAGCCATATGTTGCTGATGCTTACCAAGAT 524
QY 504 CTTTTCGCACTCCGTGGCCACTGGAGAGTGTACCTCTGTGAGCTCAGGAGCTTGG 563
Db 525 CTTTTCGCACTCCGTGGCTGTGCTTACAGAGAGTGTTCCTCGAG---TCAGAGGTTAG 581
QY 564 GGCCTGGAAGAGTGTGCAAGACAGGAGATGTGTACCGCTGAGACCTTGTCTAATCCG---- 619
Db 582 AGCTGTGAGAGGTGTAGACAGAGATAGCGTACATGAGGTGATGATGATGATG 641
QY 620 ----TTTTTATTAAACCACTCTGCGAGAGTACACAGCTGAACCAATTGTGATGCTG 677
Db 642 TTTATTTACTATATACACAGACCGGTAATAACTATTAAGCTGAAGTGTGCTGATGCTG 701
QY 678 CCTGTAAATGAGCGGAGGAGATCTTGCCTACGCCCGGAGACAGCAATGATGCTG 737
Db 702 ACCGTTATTTGATGCTGTGTGATTTTGTATGCGAGGATGATGGAATGATTCATC 761
QY 738 GAAAGCT 744
Db 762 ATTAGCT 768

RESULT 8

US-09-770-445-700
; Sequence 700, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Xu, Yang
; APPLICANT: Kameoka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; PRIOR APPLICATION NUMBER: 2001-01-26
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
; LENGTH: 829
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-700

Query Match 19.0%; Score 196.6; DB 10; Length 829;
Best Local Similarity 60.3%; Pred. No. 2.4e-44;
Matches 382; Conservative 1; Mismatches 235; Indels 15; Gaps 3;

[illegible]

```

RESULT 9
US-09-878-574-157
; Sequence 157, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:

```

```

1  APPLICANT: Byrum, Joseph R.
2  APPLICANT: La Rosa, Thomas J.
3  APPLICANT: Thompson, Michael D.
4  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
5  TITLE OF INVENTION: Plants
6  FILE REFERENCE: 38-21(15401)B
7  CURRENT APPLICATION NUMBER: US/09/878,574
8  CURRENT FILING DATE: 2001-12-21
9  PRIOR APPLICATION NUMBER: 09/333,535
10 PRIOR FILING DATE: 1999-06-14
11 NUMBER OF SEQ ID NOS: 15775
12 SEQ ID NO 157
13     LENGTH: 369
14     TYPE: DNA
15 ORGANISM: Glycine max
16 OTHER INFORMATION: Clone ID: LIB3028-054-Q1-B1-E8
17 US-09-878-574-157

```

Query Match	18.18;	Score 187.2;	DB 10;	Length 369;
Best Local Similarity	70.88;	Pred. No. 5.9e-42;		
Matches 262; Conservative	1;	Mismatches 104;	Indels 3;	Gaps 1;

QY	358	TTGCTGCTCCGGAGAGAGATCTCCCGGTACGCCATATGTGGGGAGACCAATTC	417
Db	3	TTATATGAAACCGGGGAGATCAAAACCAATATAGCCATGTCATGCGGGGATGCCAACTTC	62
QY	418	GGGTGGCGTTCCTCTCTCGGAGACACGCGCGGTGGCGCTTGGGGCCGCTGGCGCAACGGCC	477
Db	63	GGTTGGCGCTCTCCCGGAGACGCGCCCGGTGGCGCTGGGGCCCTCCGGGGCCCGGCA	122
QY	478	ATTGTTGCSAGAGGCTACGCCCGCATCTTTTTCGCAACTTCGTTGGGCCATGGAAGAGTG	537
Db	123	GTGTGTCGCGAATCTGATCTAGATCTTTCTTTCGGAACCTCGGTGGCCACCGGGAGAGTG	182
QY	538	TACCCCTGAGACTCAACGAGCTTGGGGCCCTGGGAAGAGTCGAACGACGAGGATGTGATC	597
Db	183	TATCCGCTAGAG---TCGAGAGGACGCCCTCTCGGAGAGATGCACCAACCGGCGATGTGGTG	239
QY	598	ACCGTGACCTTGTCTACTCCGTTTATTATTAACCAACTCTGGCAAGGAGTCAAACTG	657
Db	240	ACGATTGACTCGGAGAGAGCGCGCTTATCAATCAACACACCGGAAAGAGATATCGCTTG	299
QY	658	AAACCAATTGATGATGCTGGCCCTGTAAATGAGAGCGGAGAGATCTTTGCTACGACCGG	717
Db	300	AAACCGATCGGGGACGCGGGTTCAGTGTATGAGAGCGGTGGCATCTTTGGCATATGCCAGG	359
QY	718	AAGACAGGAA 727	
Db	360	AAAACCGGCA 369	

```

1  RESULT 10
2  US-09-878-574-1701
3  : Sequence 1701, Application US/09878574
4  : Patent No. US20020110548A1
5  : GENERAL INFORMATION:
6  :
7  : APPLICANT: Byrum, Joseph R.
8  :
9  : APPLICANT: La Rosa, Thomas J.
10 :
11 : APPLICANT: Thompson, Michael D.
12 :
13 : TITLE OF INVENTION: Nucleic Acid Mole
14 :
15 : FILE OF INVENTION: Plants
16 :
17 : FILE REFERENCE: 38-21(15401)B
18 :
19 : CURRENT APPLICATION NUMBER: US/09/878
20 :
21 : CURRENT FILING DATE: 2001-12-21
22 :
23 : PRIOR APPLICATION NUMBER: 09/333,535
24 :
25 : PRIOR FILING DATE: 1999-06-14
26 :
27 : NUMBER OF SEQ ID NOS: 15775
28 :
29 : SEQ ID NO 1701
30 :
31 : LENGTH: 373
32 :
33 : TYPE: DNA
34 :
35 : ORGANISM: Glycine max
36 :
37 : OTHER INFORMATION: Clone ID: LIB3028
38 :
39 : US-09-878-574-1701

```

Query Match	9.48;	Score 96.8;	DB 10;	Length 373;
Best Local Similarity	71.9%;	Pred. No. 5.5e-17;		
Matches 141;	Conservative	0;	Mismatches 52;	Indels 3;
				Gaps 1;

OY	180	CCTGTCGGCGTGTGCAGGCCGTTTCCACGGGAGTGTTCCGATGGAGGAGCAATAATCGA	239
Db	175	CGGGCGCTCCGCTCCGCTCCTTCATGAGCTCTGTCTACGTGTCTGGCGACAACTATCGA	234
OY	240	CACCGACAGATCATCCCCGCGCGACACTCATCTGTGTGGTCCCTCCAGCCGAGCGAGTA	299
Db	235	GACCGATTCAATGATTTCCCGCGGAGTAGTACCCTCGTCCCTTGGAAACCCGACAGATA	294
OY	300	CCGCGAACGTGGTTCCTTGGCCCTTCGCGGGGCTCCATCCGCGGCTTACCGAGGCCGTT	355
Db	295	CGAGAAAGCTCGGCTCCTCAAGGCCCTCGTGGGCTCC---CGGCGACTTGAAGCGCGCGCTT	351
OY	360	CGTGCCTCCGCGGTGAG	375
Db	352	CGTGCACCCCGGCGAG	367

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 02:53:39 ; Search time 1498 Seconds
(without alignments)
1168.183 Million cell updates/sec

Title: US-10-027-450-46

Perfect score: 1033
Sequence: 1 atcatggcgcgctctgtc.....aaaaaaaaaaaaaaaaa 1033

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estomv:*
5: em_estom:*
6: em_estipl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esthum:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	910.4	88.1	1085	11	AY109732
2	570.6	55.2	571	9	AY109732
3	536.8	51.9	546	9	AY109732
4	533.8	51.7	546	9	AY109732
5	526.6	51.0	556	10	AY109732
6	493	47.7	568	10	AY109732

Result No.	Score	Query Match	Length	ID	Description
7	472	45.7	601	9	AY109732
8	471.6	45.7	534	10	AY109732
9	461.2	44.6	564	10	AY109732
10	453.4	43.9	566	10	AY109732
11	453	43.9	551	12	AY109732
12	442.6	42.8	546	10	AY109732
13	435.4	42.1	841	12	AY109732
14	428.8	41.5	534	10	AY109732
15	412.6	39.9	500	12	AY109732
16	410.6	39.7	499	10	AY109732
17	409	39.6	409	13	AY109732
18	403.6	39.1	615	10	AY109732
19	388.8	37.6	691	14	AY109732
20	387.2	37.5	451	10	AY109732
21	376.6	36.5	403	9	AY109732
22	371.8	36.0	700	9	AY109732
23	371.4	36.0	393	13	AY109732
24	366	35.4	635	10	AY109732
25	365.4	35.4	619	10	AY109732
26	335	32.4	546	9	AY109732
27	330.8	32.0	557	12	AY109732
28	326.4	31.6	866	12	AY109732
29	323.4	31.3	700	9	AY109732
30	322.2	31.3	710	10	AY109732
31	310.8	30.1	373	9	AY109732
32	300.4	29.1	615	12	AY109732
33	289.8	28.1	580	14	AY109732
34	284.2	27.5	565	12	AY109732
35	283.4	27.4	369	12	AY109732
36	269	26.0	479	12	AY109732
37	258.8	25.1	834	14	AY109732
38	254.4	24.6	286	17	AY109732
39	244.4	23.7	714	9	AY109732
40	242	23.4	640	13	AY109732
41	231.6	22.4	596	10	AY109732
42	228	22.1	717	17	AY109732
43	228	22.1	746	17	AY109732
44	226.6	21.9	1344	10	AY109732
45	226.4	21.9	524	13	AY109732

ALIGNMENTS

RESULT 1
LOCUS AY109732 1085 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays CL1042_1 mRNA sequence.
ACCESSION AY109732
VERSION AY109732.1 GI:21213564
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morigane,M. and Tingey,S.V.
TITLE Maize Mapping Project/Dupont Consensus Sequences for Design of
1 (bases 1 to 1085)
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1085)
AUTHORS Coe,E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA

FEATURES
source location/Qualifiers
1..1085
/organism="Zea mays"
/db_xref="MaizeDB:629913"
/db_xref="taxon:4577"
/clone="CL1042_1"

/clone.lib="Maize Mapping Project/DuPont Consensus Library"
 /note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 203 a 303 c 295 g 239 t 45 others
 ORIGIN

Query Match 88.1%; Score 910.4; DB 11; Length 1085;
 Best Local Similarity 94.0%; Pred. No. 1.4e-136;
 Matches 941; Conservative 1; Mismatches 52; Indels 7; Gaps 2;

1 ATCATGGCGGCGCTCTGTGCGGGAGCGGGGTGTCCACGGCAGCGCTTCTAGACCCCAATC 60
 |||||||
 8 ATCATGGCGGCGCTCTGTGCGGGAGCGGGGTGTCCACGGCAGCGCTTCTAGACCCCAATC 67
 |||||||
 61 CGAGCTCCAAACGAGCGCGTTATCCGGCGCTCCAGCTCAGCTGTCATCGCCTCCAGTCA 120
 |||||||
 68 CGAGCTCCAAACGAGCGCGTTATCCGGCGCTCCAGCTCAGCTGTCATCGCCTCCAGTCA 127
 |||||||
 121 CTAAATGCCGCGCGCTGGGTTCATGTCGCCGCCGCCGCTGCTGCCGCGGCGAGC 180
 |||||||
 128 CTAAATGCCGCGCGCTGGGTTCATGTCGCCGCCGCCGCTGCTGCCGCGGCGAGC 187
 |||||||
 181 AGCTCGCGCTGTGAGCGCGTTTTCACGGCGAGTCTTGTGTGGGCGACAAATTCAGC 240
 |||||||
 188 AGCTCGCGCTGTGAGCGCGTTTTCACGGCGAGTCTTGTGTGGGCGACAAATTCAGC 247
 |||||||
 241 ACCGACCAATCATCCCGCGGAGACACATCTGTGTCGCCCTCCAAAGCGGAGACAGTAC 300
 |||||||
 248 ACCGACCAATCATCCCGCGGAGACACATCTGTGTCGCCCTCCAAAGCGGAGACAGTAC 307
 |||||||
 301 CGCAAGCTCGGTTCTTCCGCTTCGCGGGGCTCCCATCCGCGGCTTACCCGACGCCGTTTC 360
 |||||||
 308 CGCAAGCTCGGTTCTTCCGCTTCGCGGGGCTCCCATCCGCGGCTTACCCGACGCCGTTTC 367
 |||||||
 361 GTCCGCTCGGGGTGAGAGTCTCCCGCTACGCGCATATGTCGGGCGAGCAAACTTCGGG 420
 |||||||
 368 GTCCGCTCGGGGTGAGAGTCTCCCGCTACGCGCATATGTCGGGCGAGCAAACTTCGGG 427
 |||||||
 421 TGCAGTCTCTCTCGAGACACGCGCCGCTGCGGCGGCGCTGCGCGACGCGCAT 480
 |||||||
 428 TGCAGTCTCTCTCGAGACACGCGCCGCTGCGGCGGCGCTGCGCGACGCGCAT 487
 |||||||
 481 GTTGGGAGGCGCTACGCGCGCATCTTTTTCGCAACTCCGTGGCCACTGGAAGGTGAC 540
 |||||||
 488 GTTGGGAGGCGCTACGCGCGCATCTTTTTCGCAACTCCGTGGCCACTGGAAGGTGAC 547
 |||||||
 541 CCTCGGAGCTCACGAGAGTTGGGGCTGGAAGAGTGAAGAGGAGTGTGTAC 600
 |||||||
 548 CCTCGGAGCTCACGAGAGTTGGGGCTGGAAGAGTGAAGAGGAGTGTGTAC 607
 |||||||
 601 GTGACCTTGTCTAACTCCGTTTATTAACCACTCTGTGGCAAGAGTGAAGGTGAAA 660
 |||||||
 608 GTGACCTTGTCTAACTCCGTTTATTAACCACTCTGTGGCAAGAGTGAAGGTGAAA 667
 |||||||
 661 CCAATGGTGTGAGTGGGCGCTGTAATTGAGGCGGAGGAGTCTTTGGCTACGCGCCGAAG 720
 |||||||
 668 CCAATGGTGTGAGTGGGCGCTGTAATTGAGGCGGAGGAGTCTTTGGCTACGCGCCGAAG 727
 |||||||
 721 ACAGGAATGATTCGTCGAAAGCTGTGCATGAGGGAA---AGCTTATGCAAGCGAGC 775
 |||||||
 728 ACAGGAATGATTCGTCGAAAGCTGTGCATGAGGGAAAGATGAGCTTAGGAGCGAGAG 787
 |||||||
 776 CTCTCGGAGATGAAGAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 835
 |||||||
 788 CTCTCGGAGATGAAGAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 847
 |||||||
 836 GTGACGAGTGTCTAAATTAAGTTGCGGCTTACGGAATTAATGAATCAATCAATTG 895
 |||||||

Db 848 GTGACGAGTGTCTAAATTAAGTTGCGGCGCTACCGGAATTAATGAATCAATCAATTG 907
 |||||||
 QY 896 GTCTTTCACAGATCGTTTTTTTGTACTAGTA--CTGTCAATGTACTCCGCGC 953
 |||||||
 Db 908 GTCTTTCACAGATCGTTTTTTTGTACTAGTA--CTGTCAATGTACTCCGCGC 967
 |||||||
 QY 954 TGTCTACTGTTCTTATCTGTTTGAATTAAGTGTCTGTGCGCA 994
 |||||||
 Db 968 TGTCTACTGTTCTTATCTGTTTGAATTAAGTGTGTGCGCA 1008
 |||||||

RESULT 2
 AI977840/c 571 bp mRNA linear EST 27-AUG-1999
 LOCUS 496028H06.xl 496 - stressed shoot cDNA library from Wang/Bohnert
 DEFINITION lab zea mays cDNA, mRNA sequence.
 ACCESSION AI977840
 VERSION AI977840.1 GI:5791048
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 Clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 571)

REFERENCE
 AUTHORS
 TITL
 JOURNAL
 COMMENT
 Zea ESTs from various cDNA libraries sequenced at Stanford
 University
 Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 496028 row: H column: 06.

FEATURES
 source
 1..571
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone.lib="496 - stressed shoot cDNA library from
 Wang/Bohnert lab"
 /issue-type="seedling"
 /dev-stage="salt stress"
 /lab_host="E.coli XL Gold"
 /note="Organ: shoot; Vector: pBluescriptII SK(+); XR;
 Wang/Bohnert"

BASE COUNT 150 a 162 c 135 g 124 t
 ORIGIN

Query Match 55.2%; Score 570.6; DB 9; Length 571;
 Best Local Similarity 99.8%; Pred. No. 4.4e-82;
 Matches 570; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 411 CAACTGGGCGGCTCTCTGCGAGCACGCGCGCTGCGCGCTGGGCGCTGGGCGC 470
 |||||||
 Db 571 CAACTGGGCGGCTCTCTGCGAGCACGCGCGCTGCGCGCTGGGCGCTGGGCGC 512
 |||||||
 QY 471 ACGGCGCATTTGTCGAGGCGTACGCGCGCATCTTTTTCGCACTCCGTGGCCACTG 530
 |||||||
 Db 511 ACGGCGCATTTGTCGAGGCGTACGCGCGCATCTTTTTCGCACTCCGTGGCCACTG 452
 |||||||
 QY 531 AGAGGTGACCTCTGAGAGCTCACGAGACGTTGGGGCGCTGGAAGAGTGAAGAGGGA 590
 |||||||
 Db 451 AGAGGTGACCTCTGAGAGCTCACGAGACGTTGGGGCGCTGGAAGAGTGAAGAGGGA 392
 |||||||
 QY 591 TGTGTCACGCTGAGACCTTGTCTAATCGTTTATTAACCACTCTGGAAGAGTA 650
 |||||||
 Db 391 TGTGTCACGCTGAGACCTTGTCTAATCGTTTATTAACCACTCTGGAAGAGTA 332
 |||||||
 QY 651 CAAGCTGAACCAATGTGATGCTGGCCTGTAAATTGAAGCGGAGGAGTCTTCCCTA 710
 |||||||

```

|||||
Db 331 CAAAGCTGAACCAATTGGTGGATGCGCCCTGTAAATTAGAGCGGAGGAGATCTTTGGCTTA 272
QY 711 CCGCCGGAAGACAGAAATGATTCGTCGAAAGCTGTGCATGAGGAAAGCTTATGACG 770
Db 272 CCGCCGGAAGACAGAAATGATTCGTCGAAAGCTGTGCATGAGGAAAGCTTATGACG 212
QY 771 CGAGCCTCGGAGATGAAGATGAGTACCTGAGTTAGGATAGACTTACTGACCTAC 830
Db 211 CGAGCCTCGGAGATGAAGATGAGTACCTGAGTTAGGATAGACTTACTGACCTAC 152
QY 831 TTGATGTGACGGTGTCTCAAAATAAGTTGGCGCTACCGAAATATGATGATCAATCA 890
Db 151 TTGATGTGACGGTGTCTCAAAATAAGTTGGCGCTACCGAAATATGATGATCAATCA 92
QY 891 ATTGAGTCTTGTGACAGATGCTTTTGTGTACTAGTACTGTACATGATGACCT 950
Db 91 ATTGAGTCTTGTGACAGATGCTTTTGTGTACTAGTACTGTACATGATGACCT 32
QY 951 GCCTGCTACTGTTCTTATCTGTTGAAATAC 981
Db 31 GCCTGCTACTGTTCTTATCTGTTGAAATAC 1

```

```

RESULT 3
A1820332 546 bp mRNA linear EST 02-FEB-2000
LOCUS 605091F07.y1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION A1820332
VERSION A1820332.1 GI:5439411
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 546)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
JOURNAL Contact: Walbot V
COMMENT Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605091 row: F column: 07.

```

```

FEATURES
source
1..546
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site_1: EcorI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

```

```

BASE COUNT 76 a 201 c 167 g 101 t 1 others
ORIGIN

```

```

Query Match 51.9%; Score 536.6; DB 9; Length 546;
Best Local Similarity 98.7%; Pred. No. 1.2e-76;
Matches 539; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 67 CCAACGAGGCGTTATCGGCGCTCCAGCTCACTGTCAGCTCCACTCACTATAAA 126
Db 1 CCAACGAGGCGTTATCGGCGCTCCAGCTCACTGTCAGCTCCACTCACTATAAA 60

```

```

QY 127 TGCCTCCGCGCTGGGTCCATGCTCCCGCGGCGCTCTCCCGCGGCGGACACTCG 186
Db 61 TGCCTCCGCGCTGGGTCCATGCTCCCGCGGCGCTCTCCCGCGGCGGACACTCG 120
QY 187 CCGTGTACGCGTTTTCACAGGCGAGTGTCTGTGTGGGCGACAATATGACACCGAC 246
Db 121 CCGTGTACGCGTTTTCACAGGCGAGTGTCTGTGTGGGCGACAATATGACACCGAC 180
QY 247 CAGATCATCCCGCGCGAGCACTCATCTGTGCTCTCCAGCCGAGACAGTACCGCAAG 306
Db 181 CAGATCATCCCGCGCGAGCACTCATCTGTGCTCTCCAGCCGAGACAGTACCGCAAG 240
QY 307 CTGCGTCTCTTGTGCTTGTGCGGCGCTCCATCCGCGGCTACCGGACGCGCTGTGCGCT 366
Db 241 CTGCGTCTCTTGTGCTTGTGCGGCGCTCCATCCGCGGCTACCGGACGCGCTGTGCGCT 300
QY 367 CCGGCTGAGAGACTCTCCCGCTACGCCATCTTGTGCGGCGAGCACTTCCGGTCCGCT 426
Db 301 CCGGCTGAGAGACTCTCCCGCTACGCCATCTTGTGCGGCGAGCACTTCCGGTCCGCT 360
QY 427 TCCTCTCCGAGACAGCGCCGCTGCGCTTGGGCGCGCTGGCGACGCGCATTTGTGCGS 486
Db 361 TCCTCTCCGAGACAGCGCCGCTGCGCTTGGGCGCGCTGGCGACGCGCATTTGTGCGS 420
QY 487 GAGGCTACGCGCGCATCTTTTTCGCACTCCGTGGCCACTGGAGAGTGTACCTCTTG 546
Db 421 GAGGCTACGCGCGCATCTTTTTCGCACTCCGTGGCCACTGGAGAGTGTACCTCTTG 480
QY 547 GAGCTCAGGAGGTTGGGCGCTGGAGAGTGTCAAGACAGGAGTGTGTCACTGTGAGAC 606
Db 481 GAGCTCAGGAGGTTGGGCGCTGGAGAGTGTCAAGACAGGAGTGTGTCACTGTGAGAC 540
QY 607 CTTGCT 612
Db 541 CTTGCT 546

```

```

RESULT 4
A1833533 589 bp mRNA linear EST 02-FEB-2000
LOCUS 605091F07.x2 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION A1833533
VERSION A1833533.1 GI:5455843
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 589)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
JOURNAL Contact: Walbot V
COMMENT Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605091 row: F column: 07.

```

```

FEATURES
source
1..589
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site_1: EcorI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

```

```

Query Match 51.9%; Score 536.6; DB 9; Length 546;
Best Local Similarity 98.7%; Pred. No. 1.2e-76;
Matches 539; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 67 CCAACGAGGCGTTATCGGCGCTCCAGCTCACTGTCAGCTCCACTCACTATAAA 126
Db 1 CCAACGAGGCGTTATCGGCGCTCCAGCTCACTGTCAGCTCCACTCACTATAAA 60

```



```
VERSION      BE510561.1  GI:9731809
KEYWORDS     EST.
SOURCE       Zea mays.
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoideae; Andropogoneae; Zea.
              1 (bases 1 to 568)
REFERENCE    Walbot, V.
AUTHORS      Maize ESTs from various cDNA libraries sequenced at Stanford
TITLE        University
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
              Department of Biological Sciences
              Stanford University
              855 California Ave, Palo Alto, CA 94304, USA
              Tel: 650 723 2227
              Fax: 650 725 8221
              Email: walbot@stanford.edu
              Plate: 946053 row: G column: 08.
              Location/Qualifiers
FEATURES     source
              1..568
               /organism="Zea mays"
               /cultivar="OH43"
               /db_xref="taxon:4577"
               /clone_lib="946 - tassels primordialium prepared by Schmidt
               lab"
               /tissue_type="tassels"
               /dev_stage="just after the transition from vegetative to
               inflorescence development"
               /lab_host="XLOLR"
               /note="Organ: tassels; Vector: HybridZAP; Site.1: EcoRI;
               Site.2: XhoI; George Chuck dissected immature tassels
               between 1mm and 3mm. Sharon Stafield prepared the cDNA
               library in HybridZAP. Sample insert size range was 350 bp
               to 3 kb with a 1 kb average."
BASE COUNT   155 a 154 c 129 g 130 t
ORIGIN
Query Match 47.7%; Score 493; DB 10; Length 568;
Best Local Similarity 97.4%; Pred. No. 1.1e-69;
Matches 523; Conservative 1; Mismatches 6; Indels 7; Gaps 2;
OY 465 TGGCGACGCGCCATTGTTGCGAGGCGCTACGCGGCACTTTTTCGCAACTCCGTGGC 524
DB TGGCGACGCGCCATTGTTGCGAGGCGCTACGCGGCACTTTTTCGCAACTCCGTGGC 509
OY 525 CACTGAGAGGTGTACCCCTGAGGCTCAGGAGCGTTGGGCGCTGAGAGTGCAGAC 584
DB TGGCGACGCGCCATTGTTGCGAGGCGCTACGCGGCACTTTTTCGCAACTCCGTGGC 509
OY 508 CACTGAGAGGTGTACCCCTGAGGCTCAGGAGCGTTGGGCGCTGAGAGTGCAGAC 449
DB TGGCGACGCGCCATTGTTGCGAGGCGCTACGCGGCACTTTTTCGCAACTCCGTGGC 509
OY 585 AGGGATGTGTACCGGTGACCTGCTAATCCGTTTATTACCAACACCTCTGGCAA 644
DB AGGGATGTGTACCGGTGACCTGCTAATCCGTTTATTACCAACACCTCTGGCAA 389
OY 645 GGAGTCAAGCTGAACCAATGCGATGCTGCGCCCTTAATGAGGCGGAGGATCTT 704
DB GGAGTCAAGCTGAACCAATGCGATGCTGCGCCCTTAATGAGGCGGAGGATCTT 329
OY 388 GGAGTCAAGCTGAACCAATGCGATGCTGCGCCCTTAATGAGGCGGAGGATCTT 329
DB GGAGTCAAGCTGAACCAATGCGATGCTGCGCCCTTAATGAGGCGGAGGATCTT 329
OY 705 TGGCTACGCCCGGAGACAGATATGCGTGAAGAGCTGCTGATGAGGAA-----A 759
DB TGGCTACGCCCGGAGACAGATATGCGTGAAGAGCTGCTGATGAGGAA-----A 269
OY 328 TGGCTACGCCCGGAGACAGATATGCGTGAAGAGCTGCTGATGAGGAA-----A 269
DB TGGCTACGCCCGGAGACAGATATGCGTGAAGAGCTGCTGATGAGGAA-----A 269
OY 760 GCTTATGACGCGAGCGCTTGGGAGATGAAGAACTAAGCTGAGTAAAGATT 819
DB GCTTATGACGCGAGCGCTTGGGAGATGAAGAACTAAGCTGAGTAAAGATT 209
OY 268 GCTTATGACGCGAGCGCTTGGGAGATGAAGAACTAAGCTGAGTAAAGATT 209
DB GCTTATGACGCGAGCGCTTGGGAGATGAAGAACTAAGCTGAGTAAAGATT 209
OY 820 ACTGACACTACTGATGTGAGAGGTGTCTCAAAATAGTTGGGCGCTACCAAAATTGTA 879
DB ACTGACACTACTGATGTGAGAGGTGTCTCAAAATAGTTGGGCGCTACCAAAATTGTA 149
OY 208 ACTGACACTACTGATGTGAGAGGTGTCTCAAAATAGTTGGGCGCTACCAAAATTGTA 149
DB ACTGACACTACTGATGTGAGAGGTGTCTCAAAATAGTTGGGCGCTACCAAAATTGTA 149
OY 880 TGAATCAATCAATTGGTCTTTGTTCACAGATGTTTTTTT--TTTGTACTGTACTGTA 937
DB TGAATCAATCAATTGGTCTTTGTTCACAGATGTTTTTTT--TTTGTACTGTACTGTA 937
```

```
DB 148 TGAATCAATCAATTGGTCTTTGTTCACAGATGTTTTTTTGTACTGTACTGTA 89
OY 938 CAATTACTCTGCGCTGCTACTGTTCTTATCTGTTGAATACTGCTGTGGCA 994
DB CAATTACTCTGCGCTGCTACTGTTCTTATCTGTTGAATACTGCTGTGGCA 32
RESULT 7
A1677354/C 601 bp mRNA linear EST 02-FEB-2000
LOCUS 605053H06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION A1677354
VERSION A1677354.1 GI:4886234
KEYWORDS Zea mays.
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoideae; Andropogoneae; Zea.
              1 (bases 1 to 601)
REFERENCE    Walbot, V.
AUTHORS      Maize ESTs from various cDNA libraries sequenced at Stanford
TITLE        University
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
              Department of Biological Sciences
              Stanford University
              855 California Ave, Palo Alto, CA 94304, USA
              Tel: 650 723 2227
              Fax: 650 725 8221
              Email: walbot@stanford.edu
              Plate: 605053 row: H column: 06.
              Location/Qualifiers
FEATURES     source
              1..601
               /organism="Zea mays"
               /cultivar="Ohio43"
               /db_xref="taxon:4577"
               /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
               /tissue_type="nucellar, embryo, and endosperm"
               /dev_stage="10-14 days post-pollination"
               /lab_host="DH5(alpha)"
               /note="Organ: Kernel; Vector: PAD-GAL4-2'; Site.1: EcoRI;
               Site.2: XhoI; Kernel endosperm cDNA library from Schmidt
               lab"
BASE COUNT   172 a 161 c 129 g 139 t
ORIGIN
Query Match 45.7%; Score 472; DB 9; Length 601;
Best Local Similarity 96.9%; Pred. No. 2.6e-66;
Matches 504; Conservative 1; Mismatches 6; Indels 9; Gaps 2;
OY 484 GCGAGGCGGTACGCGGCACTTTTTCGCAACCGCGGCGCTGAGAGGTACCT 543
DB GCGAGGCGGTACGCGGCACTTTTTCGCAACCGCGGCGCTGAGAGGTACCT 542
OY 544 CTGAGCTCAGCGAGCTTGGGCGCTGGAAGAGTGCAGACAGGAGGATGTGTACCGTG 603
DB CTGAGCTCAGCGAGCTTGGGCGCTGGAAGAGTGCAGACAGGAGGATGTGTACCGTG 482
OY 541 CTGAGCTCAGCGAGCTTGGGCGCTGGAAGAGTGCAGACAGGAGGATGTGTACCGTG 482
DB CTGAGCTCAGCGAGCTTGGGCGCTGGAAGAGTGCAGACAGGAGGATGTGTACCGTG 482
OY 604 GACCTTGCTAATCCGCTTTTATTACCAACACCTCTGCAAGAGTCAAGCTGAACCA 663
DB GACCTTGCTAATCCGCTTTTATTACCAACACCTCTGCAAGAGTCAAGCTGAACCA 422
OY 481 GACCTTGCTAATCCGCTTTTATTACCAACACCTCTGCAAGAGTCAAGCTGAACCA 422
DB GACCTTGCTAATCCGCTTTTATTACCAACACCTCTGCAAGAGTCAAGCTGAACCA 422
OY 664 ATTGGTATGTGCGCCCTGTAATTGAGCGGAGGAGATCTTGGCTACGCCCGGAACACA 723
DB ATTGGTATGTGCGCCCTGTAATTGAGCGGAGGAGATCTTGGCTACGCCCGGAACACA 362
OY 421 ATTGGTATGTGCGCCCTGTAATTGAGCGGAGGAGATCTTGGCTACGCCCGGAACACA 362
DB ATTGGTATGTGCGCCCTGTAATTGAGCGGAGGAGATCTTGGCTACGCCCGGAACACA 362
OY 724 GGAATGATGGCTGGAAGAGCTGTCATGAGAGGAA-----AGCTTATGACGCGAGGCTC 778
DB GGAATGATGGCTGGAAGAGCTGTCATGAGAGGAA-----AGCTTATGACGCGAGGCTC 302
OY 361 GGAATGATGGCTGGAAGAGCTGTCATGAGAGGAA-----AGCTTATGACGCGAGGCTC 302
DB GGAATGATGGCTGGAAGAGCTGTCATGAGAGGAA-----AGCTTATGACGCGAGGCTC 302
OY 779 TGGGAGATGAAGAAGTAAGCTGAGTGAAGTAAAGATTACTGACACTTGTATGTC 838
DB TGGGAGATGAAGAAGTAAGCTGAGTGAAGTAAAGATTACTGACACTTGTATGTC 838
```

|||||
Db 301 TCCGAGATGAGAGTACGAGTTAGAGTACGACCTACTGATGTC 242
OY 839 GACGGTGTCAAAATAGTTGCGCCCTACCGAAATATGATGATCAATTTGTC 898
Db 241 GACGGTGTCAAAATAGTTGCGCCCTACCGAAATATGATGATCAATTTGTC 182
OY 899 TTGGACAGATCGTTTTT---TTGGTACTAGTGTGCAATTGATGCGGCC 954
Db 181 TTGGACAGATCGTTTTTTTTGTGTGTTACTGATGATCAATTTGATCCT 122
OY 935 GCTACTGTCTTATCTTTTGAATACGCTCTGTGCCA 994
Db 121 GCTACTGTCTTATCTTTTGAATACGCTCTGTGCCA 82
RESULT 8
LOCUS BE597022 534 bp mRNA linear EST 18-AUG-2000
DEFINITION P11_60_D05.b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION BE597022
VERSION BE597022.1 GI:9852095
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 534)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 533
POLYA-No.
FEATURES
Source location/Qualifiers
1..534
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_id="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
Vector: pluescript II from Lambda Zap II; Site.1: XhoI;
Site.2: EcoRI; Two-week-old sorghum plants (B7X 623
cultivar) were infected with pathogen (isolate PM44J of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from polyA
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."
BASE COUNT 95 a 173 c 167 g 99 t
ORIGIN
Query Match 45.7%; Score 471.6; DB 10; Length 534;

Best Local Similarity 93.2%; Pred. No. 3,1e-66;
Matches 492; Conservative 1; Mismatches 35; Indels 0; Gaps 0;
OY 226 GCGCAATATGACACCGACAGATATCCCGCGAGCACTGCTGTCCTCC 285
Db 7 GCGCAATATGACACCGACAGATATCCCGCGAGCACTGCTGTCCTCC 66
OY 286 AAGCGGAGAGTACCGCAAGCTCGTTCCTGCGCGGCGGCTCCATCGCGGCC 345
Db 67 AAGCTGACAGATACCGCAAGCTCGTTCCTGCGCGGCGGCTCCATCGCGGCC 126
OY 346 TACCGAGCGCGTTCGTCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 405
Db 127 TACCGAGCGCGTTCGTCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 186
OY 406 GGAGCGACCTTCGCGTTCGTCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 465
Db 187 GGCGCAATTCGCGTTCGTCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 246
OY 466 GGCGCAATTCGCGTTCGTCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 525
Db 247 GGAGCGCGCGCGTTCGTCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 306
OY 526 ACTGAGAGGTGTACCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 585
Db 307 ACTGAGAGGTGTACCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 366
OY 586 GGGAGTGTGTACCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 645
Db 367 GGGAGTGTGTACCGCTCGGAGAGTCCCTCCGATGAGCAATTTGCGC 426
OY 646 GAGTACAGCTGGAACCAATTTGATGATGCGCGCTGATGAGCGGAGGATCTT 705
Db 427 GAGTACAGCTGGAACCAATTTGATGATGCGCGCTGATGAGCGGAGGATCTT 486
OY 706 GCTACGCGCGGAGAGCAAGATGATGCGTGAAGCTGTCATGA 753
Db 487 GCTACGCGCGGAGAGCAAGATGATGCGTGAAGCTGTCATGA 534
RESULT 9
LOCUS BE360232 564 bp mRNA linear EST 20-JUL-2000
DEFINITION Dg1_62_G01.b1_A002 Dark grown 1 (Dg1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE360232
VERSION BE360232.1 GI:9301789
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 564)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 543
POLYA-No.
FEATURES
Source location/Qualifiers
1..564

QY	358	TTGTGTCGTC	CCGGTATAGAGATCTCCCGTACGCATCTATTGTGGCGGAGCAACTTC	417
Db	384	TTGTGTCGCCCGGGTATAGAGATCCACCCGCTACGCCGCTGCTGCGGTGGGCAACTTC	443	
QY	418	GGGTGGGGTTTCTCTGTGGCAGCAGCCGCGCTGGGGGGCGCTGGCGCAGCGGC	477	
Db	444	GGGTGGGGCTCTCTCCCGGAGCAGCGCGCGCTGCGCTTGGGGGGGCTTGGAGCGCGCGGC	503	
QY	478	ATTGTGTCGAGAGGCTACGCGCGCATTTTTCGGCAACTCG	520	
Db	504	GTCCTCGCGAGGGCTACGCGCGCATCTTCTCGCAACTCG	546	

RESULT 13					
LOCUS	BG309960				
DEFINITION	BG309960	841 bp	mrna	linear	EST 22-OCT-2001
ACCESSION	HYSMC0015H22f	Hordeum vulgare seedling shoot EST library			
VERSION	HYSCDA0003	(E) isolated and unstresssed) Hordeum vulgare cDNA clone			
KEYWORDS	HYSMC0015H22f,	mRNA sequence.			
SOURCE	BG309960				
ORGANISM	BG309960.1	GI:13110807			
	EST.				
	Hordeum vulgare.				
	Hordeum vulgare.				

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (Pages 1 to 84)	Wang, R., Close, T. J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D. W., Fenton, R. D., Oates, R. and Main, D.	Development of a genetically and physically anchored EST resource for barley genomics: Korex unstressed seedling shoot cDNA library	Unpublished (2001)

JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Total hg bases = 553
 seq primer: AATTACCTCAGTAAAGG
 High quality sequence stop: 712.
FEATURES Location/Qualifiers
SOURCE 1. .841

`"organism": "Hordeum vulgare",
 "cultivar": "Morex",
 "db_xref": "taxon:4513",
 "clone": "HVSME0015H22F",
 "clone_1ib": "Hordeum vulgare seedling shoot EST library
HVCNDA0003 (Etolated and unstressed)",
 "tissue_type": "Seedling shoot",
 "lab_host": "R10121",
 "note": "Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper
 with water, mystatin and cefotaxime in covered
 crystallization dishes. Five-day old seedling shoots were
 then harvested, total RNA was prepared, poly(A) RNA was
 purified, one primary unamplified cDNA library was made,
 and 1 million p1u were in vivo excised to give plasmid
 SK(-) cDNA phageids. These steps were performed in the
 TSC close laboratory at the University of California,
 Riverside (Choi, Close, Fenton). Phageids were plated and
 picked at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Fisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Fisch, Henry, Simmons, Oatsee,
 Rambo, Main). The sequence has been trimmed to remove
 vector sequence and contains a minimum of 100 bases of
 phred value 20 or above. For more details on library
 preparation and sequence analysis see`

Query Match	42.1%	Score 435.4	DB 12	Length 841
Best Local Similarity	77.0%	Pred. No. 1.5e-60		
Matches 356	Conservative	1	Mismatches 151	Indels 15
				Gaps 2
BASE COUNT	136 a	308 c	247 g	146 t
ORIGIN	4 others			

BASE COUNT	136 a	308 c	247 g	146 t	4 others
Query Match	42.1%	Score 435.4	DB 12	Length 841	
Best Local Similarity	77.0%	Pred. No. 1.5e-60			
Matches 538	Conservative	1	Mismatches 151	Indels 15	Gaps 2
OY	22	GGAGGCGGGGTGTCACGAGCGGCTTCAGCCCAATCCGAGCTCCACACGCGGTT	81		
DB	93	GGCGGACGAGCGAGCGAGCGAGCGGTGCTAGGCCCCCGGCGGAGCGCCACAGAGCCCG	152		
OY	82	ATCCGGGCGCTCCAGCTCACTACCTGTATCGCCCTCCACTACTAAATATCCGCGCGCTGG	141		
DB	153	ATTATGGCGCGCCCTCGCCACCCCGCTACCG--CTCGGTAATATCCACCGCTCTCG	209		
OY	142	TCCATCTCTCCCGCGCGCGCTGCTGCTCGCGGCGGAGCAGCTCGCGCTGTACGCGTT	201		
DB	210	CCCTTACCGCGCGCGCGCGCGAGCGA-----CTCGCGCTGTGCGCGCG	257		
OY	202	TTCCACGGCGAGTCTTCTGTGTGGGCGGACATATCGACACCCGACGATATCCCCCG	261		
DB	258	TTTCACGCGGAGTCTTCTGTGTGGGCGGACACATCGACACCGACGATATCTCCGCG	317		
OY	262	GAGACCTCACTGTGGGCGCTTCACGCGGAGCGAGTACCGCAGACTCGGTTCTTCGCG	321		
DB	318	GAGACCTCACTGTGGGCGCTTCACGCGGAGCGAGTACCGCAGACTCGGTTCTTCGCG	377		
OY	322	TTCCGGGGGCTCCCATCCGCGGCTACCGAGCGCGTTCGCTCGCTCCGCGGTGAGAGTC	381		
DB	378	TTCTGTGGGCTCCCTCCCTCCGCGGCTACCGGATTTCCCTTTGTGGCCCGCGGAGGATCG	437		
OY	382	TCCGCTACGCAATCTGTGGCGGAGCGCACTTCGCGGTGCGCTCTCTCGCGAGCAC	441		
DB	438	TCAGCTACGCGCTGCACTACCGCGGCGCGCACTTCGAGATCGGCTCTCCCGGAGCAC	497		
OY	442	GGCGCCGCTCGGCTTGGGGCGCGCTGCGGCGACGCGCATTTGTCGAGAGGCTACGCGCG	501		
DB	498	GGCGCCGCTCGGCTTGGGGCGCGCGCGCGCGCGCGCGCGCTTGGGCGGAGGCTACGCGCG	557		
OY	502	ATCTTTTTCGCAACTCCGCGGCGCACTGAGAGAGTGAACCTCTGAGAGCTCAAGSGATT	561		
DB	558	ATCTTTTTCGCAACTCCGCGGCGCACTGAGAGAGTGAACCTCTGAGAGCTCAAGSGATT	617		
OY	562	GGGCGCTCGAAGAGTGCACAGACAGGGGATGTGTACCGCTGAGACTTGTCTACTCCGTT	621		
DB	618	GGTCCGACGAGGAGTGCATAGACCCAGATGTGTGCACAGTTCGATGATCTCGCTC	677		
OY	622	TTTATTATACCACTCTCGCAGAGAGTATCAAGCTGGAACCAATTTGATGTGTCGCCCT	681		
DB	678	TTGATTTACCACTATTTTCGATATTCGATATTCGATATTCGATATTCGATATTCGATAT	737		
OY	682	GTAATTGAGCGGGAGGAGTCTTGTGCTTACGCGCGGAGAGACAGATGATGTGCTGAAA	741		
DB	738	GNTATTGAGCGCGGGAGGAGTCTTGTGCTTACGCGCGGAGAGACTGATGATGATGCTCCAA	797		
OY	742	GCTGC 746			
DB	798	TCTGC 802			
RESULT 14					
AM747153					
LOCUS	AM747153	534 bp	mrna	linear	EST 19-JUL-2000
DEFINITION	WS1.66.G07.b1.A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,				
ACCESSION	AM747153				
VERSION	AM747153.1	GI:7660891			


```

Db      200 GGTGCTCAAAATAGTTGCGGCTACCGAAATATGATGAATCATCAATTTGCTTT 141
Qy      902 GTCACAGATCGTTTTT---TTTGTACTAGTACTGTACAAATGTACTCTGCTGCT 957
Db      140 GTCACAGATCGTTTTT---TTTGTACTAGTACTGTACAAATGTACTCTGCTGCT 81
Qy      958 ACTGTCTTATCTGTTGAATAACTGCTCTGTGCCA 994
Db      80 ACTGCTTATCTGTTGAATAACTGCTGTGCCA 44

```

Search completed: April 3, 2003, 04:15:21
 Job time : 1508 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 02:59:38 ; Search time 75 Seconds
(without alignments)
4223.961 Million cell updates/sec

Title: US-10-027-450-46
Perfect score: 1033
Sequence: 1 atcatgagcgagcgtctgtc.....aaaaaaaaaaaaaaaa 1033

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032.6	100.0	1033	4	US-09-173-300-46
2	961.2	93.0	995	4	US-09-173-300-52
3	462.2	44.7	1112	4	US-09-173-300-48
4	291.2	28.2	1107	4	US-09-173-300-50
5	57.2	5.5	4403765	4	US-09-103-840A-2
6	57	5.5	1466	4	US-08-984-919A-10
7	57	5.5	1466	4	US-08-984-919A-12
8	57	5.5	1472	4	US-08-781-420-10
9	57	5.5	1472	4	US-08-781-420-12
10	57	5.5	1472	4	US-08-874-102-10
11	57	5.5	1472	4	US-08-874-102-12
12	57	5.5	1472	4	US-09-006-595A-10
13	57	5.5	1472	4	US-09-006-595A-12
14	57	5.5	1875	4	US-08-984-919A-46
15	57	5.5	1875	4	US-08-984-919A-48
16	57	5.5	1881	4	US-08-874-102-46
17	57	5.5	1881	4	US-08-874-102-48
18	55.4	5.4	1772	2	US-08-960-022-13
19	53.4	5.2	985	4	US-09-056-556-182
20	53.4	5.2	985	4	US-09-072-596-177
21	52	5.0	991	3	US-08-924-747-25
22	52	5.0	991	4	US-09-247-373B-25
23	52	5.0	991	4	US-09-296-715-25
24	51.8	5.0	1359	4	US-09-387-574-11
25	51.8	5.0	1359	4	US-09-387-574-11
26	51.6	5.0	3320	4	US-09-394-200-1
27	51.2	5.0	1534	1	US-08-300-903A-6

28	51.2	5.0	3080	4	US-09-099-041A-25	Sequence 25, Appl
29	51.2	5.0	3080	4	US-09-245-281-25	Sequence 25, Appl
30	51.2	5.0	3080	4	US-09-207-359B-25	Sequence 25, Appl
31	50.2	4.9	3437	3	US-08-860-339-17	Sequence 17, Appl
32	50.2	4.8	6476	4	US-09-127-670-5	Sequence 5, Appl
33	49.4	4.8	44377	2	US-08-804-227C-7	Sequence 7, Appl
34	49.4	4.8	44377	2	US-08-804-198-1	Sequence 1, Appl
35	49.2	4.8	1474	4	US-08-821-994-64	Sequence 64, Appl
36	48.8	4.7	1313	4	US-09-149-476-112	Sequence 112, Appl
37	48.8	4.7	1872	4	US-09-801-052-1	Sequence 1, Appl
38	48.8	4.7	2836	3	US-08-747-221B-24	Sequence 24, Appl
39	48.8	4.7	2836	3	US-08-747-221B-26	Sequence 26, Appl
40	48.8	4.7	2836	4	US-09-005-051-24	Sequence 24, Appl
41	48.8	4.7	2836	4	US-09-005-051-26	Sequence 26, Appl
42	48.8	4.7	3024	6	5284931-1	Patent No. 5284931
43	48.6	4.7	1098	3	US-09-248-335-35	Sequence 35, Appl
44	48.6	4.7	8920	2	US-08-446-855A-1	Sequence 1, Appl
45	48.6	4.7	8920	4	US-09-150-741-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-09-173-300-46
; Sequence 46, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Canoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 46
; LENGTH: 1033
; TYPE: DNA
; ORGANISM: Zea mays
US-09-173-300-46

Query Match      100.0%  Score 1032.6; DB 4; Length 1033;
Best Local Similarity 100.0%; Pred. No. 4.7e-226;
Matches 1033; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCATGCGGCGGCTGTGCGGGAGCGGCGGTCTACAGGCGGCTTCTAGCCCAATC 60
    |||||
Db 1 ATCATGCGGCGGCTGTGCGGGAGCGGCGGTCTACAGGCGGCTTCTAGCCCAATC 60

OY 61 CGAGCTCCAAACGAGCGGCTTATTCGCGGCTCCAGCTACCTGTATCGCTCCACTCA 120
    |||||
Db 61 CGAGCTCCAAACGAGCGGCTTATTCGCGGCTCCAGCTACCTGTATCGCTCCACTCA 120

OY 121 CTAAATGCGCGGCGGCTGTGCGGAGCATGTCGCCGCGGCGGTGCGGCGGCGGAGC 180
    |||||
Db 121 CTAAATGCGCGGCGGCTGTGCGGAGCATGTCGCCGCGGCGGTGCGGCGGCGGAGC 180

OY 181 AGCTGCGCGTGTGAGCGGCTTTCACAGGCGAGTCTTGCTGTGTGGCGGACATATCGAC 240
    |||||
Db 181 AGCTGCGCGTGTGAGCGGCTTTCACAGGCGAGTCTTGCTGTGTGGCGGACATATCGAC 240

OY 241 ACCGACAGATCATTCGCCCGGAGCACTCATCTGTGCGGCGGCGGAGGAGATAC 300
    |||||
Db 241 ACCGACAGATCATTCGCCCGGAGCACTCATCTGTGCGGCGGCGGAGGAGATAC 300

OY 301 CGCAGCTGGTTCCTTGGCTTTCGCGGCGGCTCCATCCGCGGCTTACGAGCCGCTTC 360
    |||||
Db 301 CGCAGCTGGTTCCTTGGCTTTCGCGGCGGCTCCATCCGCGGCTTACGAGCCGCTTC 360
```

Db	301	CGAAGCTCGGTTCTTCGCTTCGCTTCGCGGGGGCTCCCATCCGCGGGCTACCCGAGCGGCTTC	360
Qy	361	GTGCGCTCCGGGTGAGGAGTCCCTCCCGCTACGCCATCTTGTGCGGGAGCCAACTTGGG	420
Db	361	GTGCGCTCCGGGTGAGGAGTCCCTCCCGCTACGCCATCTTGTGCGGGAGCCAACTTGGG	420
Qy	421	TGGCGTTCTCTCGCGAGCACGCGCCCGTGGCGCTTGGGGCCGGTGGCGAGCGGCATT	480
Db	421	TGGCGTTCTCTCGCGAGCACGCGCCCGTGGCGCTTGGGGCCGGTGGCGAGCGGCATT	480
Qy	481	GTTCGSGAGGCGTACGCGCGCATCTTTTTCGCAACTCGGTGGCCATGGAGAGGTGTAC	540
Db	481	GTTCGSGAGGCGTACGCGCGCATCTTTTTCGCAACTCGGTGGCCATGGAGAGGTGTAC	540
Qy	541	CCTCTGAGAGCTCAGGAGAGTGGGGCCCGAAGAGAGTGCAAGACAGGGGAGTGGGTACC	600
Db	541	CCTCTGAGAGCTCAGGAGAGTGGGGCCCGAAGAGAGTGCAAGACAGGGGAGTGGGTACC	600
Qy	601	GTGACACTTGGTAACTCCGTTTTTATTAAACACACACTCTGGCAAGAGTACAAAGCTGAAA	660
Db	601	GTGACACTTGGTAACTCCGTTTTTATTAAACACACACTCTGGCAAGAGTACAAAGCTGAAA	660
Qy	661	CCAATTGGTATGCTGGGCCCTGTATTTAGAGCGGGAGGAGATCTTTGGCTACGCCGGAG	720
Db	661	CCAATTGGTATGCTGGGCCCTGTATTTAGAGCGGGAGGAGATCTTTGGCTACGCCGGAG	720
Qy	721	ACAGAAATGATTCGCTCCAAAGCTCTCATGAGAGGAAAGTATTGACAGCCAGCCTCTG	780
Db	721	ACAGAAATGATTCGCTCCAAAGCTCTCATGAGAGGAAAGTATTGACAGCCAGCCTCTG	780
Qy	781	CGGAGATGAAGAATGAGCTGGAGTTAGAGACTAGAGTTACTGCACTACTTGAATGTCGA	840
Db	781	CGGAGATGAAGAATGAGCTGGAGTTAGAGACTAGAGTTACTGCACTACTTGAATGTCGA	840
Qy	841	CGGTGTCCTCAAAATGAGTTGGGGCCCTACCGAAATATATGATCAATCAATTTGGCTCT	900
Db	841	CGGTGTCCTCAAAATGAGTTGGGGCCCTACCGAAATATATGATCAATCAATTTGGCTCT	900
Qy	901	TGTCAAGATTCCTTTTTTTTTTTTGTACTAGTACTGTACAAATTTGACTCTGCTGCTACT	960
Db	901	TGTCAAGATTCCTTTTTTTTTTTTGTACTAGTACTGTACAAATTTGACTCTGCTGCTACT	960
Qy	961	GTTCTTATCTGTTTGAAATACGCTCTGTGGCCAAAAAATTTTTTTTTTTTTTTTTTTT	1020
Db	961	GTTCTTATCTGTTTGAAATACGCTCTGTGGCCAAAAAATTTTTTTTTTTTTTTTTTTT	1020
Qy	1021	AAAAAAAAAAAAA 1033	
Db	1021	AAAAAAAAAAAAA 1033	
RESULT 2			
US-09-173-300-52			
Sequence 52, Application US/09173300			
Patent No. 6451581			
GENERAL INFORMATION:			
APPLICANT: Falco, Saverio Carl			
APPLICANT: Hitz, William D.			
APPLICANT: Kinney, Anthony J.			
APPLICANT: Cahoon, Rebecca E.			
APPLICANT: Rafalski, J. Antoni			
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES			
FILE REFERENCE: BR-1126			
CURRENT APPLICATION NUMBER: US/09/173,300			
CURRENT FILING DATE: 1998-10-15			
EARLIER APPLICATION NUMBER: 60/063,423			
EARLIER FILING DATE: 1997 October 28			
NUMBER OF SEQ ID NOS: 54			
SOFTWARE: Microsoft Word Version 7.0A			
SEQ ID NO 52			
LENGTH: 995			
TYPE: DNA			
ORGANISM: Trifolium aestivum			

US-09-173 -300 -52									
Query Match	93.0%;	Score 961.2;	DB 4;	Length 995;					
Best Local Similarity	99.0%;	Pred. No. 8e-210;							
Matches 978;	Conservative	1;	Mismatches	4;	Indels	5;	Gaps	1;	
Qy	26	CGGCGGTGCCACGGCAGCAGCTTCTAGGCCCAATCCAGATCTCAACCAAGCGGCTTAATCC	85						
Db	8	CGGCGGTGCCACGGCAGCAGCTTCTAGGCCCAATCCAGATCTCAACCAAGCGGCTTAATCC	67						
Qy	86	GGCGCTCCAGCTCACTGTCATGCGCTCCACTACTATAATGCGCGCGGCTGGGTCCA	145						
Db	68	GGCGCTCCAGCTCACTGTCATGCGCTCCACTACTATAATGCGCGCGGCTGGGTCCA	127						
Qy	146	TGCTGCCCGCGCGCGCTCTGCGCGGGGGGAGAGCTGCGCTGCTCAAGCCGTTTTTC	205						
Db	128	TGCTGCCCGCGCGCGCTCTGCGCGGGGGGAGAGCTGCGCTGCTCAAGCCGTTTTTC	187						
Qy	206	ACGGCGAGTGTCTTGCTGGGCGCAATATGACACCCAGCATATCCCGCCGAGC	265						
Db	188	ACGGCGAGTGTCTTGCTGGGCGCAATATGACACCCAGCATATCCCGCCGAGC	247						
Qy	266	ACCTCACTTGCTGCTCCATCAAGCGCGAGATCCGACCTGGTCTCTGCTTCG	325						
Db	248	ACCTCACTTGCTGCTCCATCAAGCGCGAGATCCGACCTGGTCTCTGCTTCG	307						
Qy	326	CGGGGCTCCCATCCGCGGCTACCCGAGCGCTTCGCTCCGGGTGAGAGTCTTCC	385						
Db	308	CGGGGCTCCCATCCGCGGCTACCCGAGCGCTTCGCTCCGGGTGAGAGTCTTCC	367						
Qy	386	GCTACGCCATCATTTGTGCGCGAGGACCACTTGGGTGGTTCCTTCGCGAGCACGGC	445						
Db	368	GCTACGCCATCATTTGTGCGCGAGGACCACTTGGGTGGTTCCTTCGCGAGCACGGC	427						
Qy	446	CGCTCCGCTTGAGGGCGCTGCGACGCGCATTTGTGSGAGGGCTACGCGCGCATCT	505						
Db	428	CGCTCCGCTTGAGGGCGCTGCGACGCGCATTTGTGSGAGGGCTACGCGCGCATCT	487						
Qy	506	TTTTTGGCAACTCGTGGCCACTGGAGAGGTGTACCTCTGGAGCTCACGAGCGTGGGG	565						
Db	488	TTTTTGGCAACTCGTGGCCACTGGAGAGGTGTACCTCTGGAGCTCACGAGCGTGGGG	547						
Qy	566	CTGGGAAGAGTGCAGACAGGGGAGTGTACCGGTGACCTTCCTTAATCCGTTTTTAA	625						
Db	548	CTGGGAAGAGTGCAGACAGGGGAGTGTACCGGTGACCTTCCTTAATCCGTTTTTAA	607						
Qy	626	TTTAACACACCTCTGGCAAGAGATACAAAGCTAATAACCAATTGGTATGCTGGCCCTGTA	685						
Db	608	TTTAACACACCTCTGGCAAGAGATACAAAGCTAATAACCAATTGGTATGCTGGCCCTGTA	667						
Qy	686	TTTAAAGCGGGAGGATCTTGGCTACGCCCGGGAAGAAGATGATGTTGGTGGAAAGGTG	745						
Db	668	TTTAAAGCGGGAGGATCTTGGCTACGCCCGGGAAGAAGATGATGTTGGTGGAAAGGTG	727						
Qy	746	CTGCGATGAGGGAA-----AGCTTATGCACCGGAGCTCTGCGGAGATGAAGAATAGCT	800						
Db	728	CTGCGATGAGGGAAAGATCAAGCTTATGCAGCGGAGCTCTGCGGAGATGAAGAATAGCT	787						
Qy	801	GGAGTTAGAGCTAAGAGTTATGCACTTATGATGATGATGATGATGATGATGATGATG	860						
Db	788	GGAGTTAGAGCTAAGAGTTATGCACTTATGATGATGATGATGATGATGATGATGATG	847						
Qy	861	CGGCTACCGGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG	920						
Db	848	CGGCTACCGGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG	907						
Qy	921	TGTTACTAGTACTTGTACAAATTTGACTCGCTGCTACTGTTCTTATCTGTTGAATA	980						
Db	908	TGTTACTAGTACTTGTACAAATTTGACTCGCTGCTACTGTTCTTATCTGTTGAATA	967						
Qy	981	CTGCTCTGTTGCCAAAAAATTTTTTTTTT	1008						
Db	968	CTGCTCTGTTGCCATTTAAAAAATTTTTTTTTT	995						

```
RESULT 3
US-09-173-300-48
; Sequence 48, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173.300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 48
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-173-300-48

Query Match
Best Local Similarity 44.7%; Score 462.2; DB 4; Length 1112;
Matches 570; Conservative 1; Mismatches 149; Indels 9; Gaps 1;

27 GGGGCTGTCACGCGAGCCCTTCTAGCCCAATCCGACCGACGCGGCTTATCCG 86
Db 72 GGGGGGGGGGGGAGACAGAGTTCTGACACCGGTCCACGCGCTTGAGAGCTTCCGCG 131
QY 87 GGGCTCCAGCTACCTGTCTATGCTCTCACTAAATGCGCGCGCTGATCAT 146
Db 132 CCGCACTGCTGGGCTATGCGCGCGCGCGCGCTGAAATGCGACAGTCTGCTT 191
QY 147 C-----GTCCCGCGGCGCTGCTGCGCGGCGGCGAGCGTGGCTGTCAGC 197
Db 192 GACCGCGGTGTCGCGCGGCTGCGCGCTGCGCGCGGCGGAGCTGACGTGCGCG 251
QY 198 GGTTCACAGCGAGTCTTCTGCTGCGGCGACAAATATGACACCGACAGATATCC 257
Db 252 CGATTTCACAGCGAGTCTTCTGCTGCGGCGATATGACACCGACAGATATCC 311
QY 258 CGCGGACGACCTACTCTGCTGCTCCCAAGCGGAGATACGCAAGCTGCGTCTT 317
Db 312 GGGCGAGACCTGACCTGCTGCTCCCAAGCGGAGATACGCAAGCTGCGTCTT 371
QY 318 CGCTTCGCGGCGCTCCATCCGCGGCTACCCGAGCGCTTCTGCTGCGGATGAGA 377
Db 372 CGCTTCGCTGCGGCTCCATCCGCGGCTACCCGAGCGCTTCTGCTGCGGATGAGA 431
QY 378 GTCCTCCGCTAGCGCATATTGTGCGGAGCAACTTGGGATGCGGTTCTCTCGGGA 437
Db 432 GACACACCGCTAGCGCATATTGTGCGGAGCAACTTGGGATGCGGTTCTCTCGGGA 491
QY 438 GCAAGCGCGCGCTGCGCTTGGGCGCTGCGGAGCGGCACTTGTGCGGAGGCTAGC 497
Db 492 GCAAGCGCGCGCTGCGCTTGGGCGCGCGCGCGCGCGCTGCTGCGGAGGCTAGC 551
QY 498 GCGCATCTTTTTCGCACTGCTGCGGCACTGAGAGATACCTCTGCGGAGCTAGCGGA 557
Db 552 GCGCATCTTCTTCCGCACTCCGTGCGGCACTGAGAGATACCTCTGCGGAGCTAGCGGA 611
QY 558 GGTGGGCGCTGAGAGAGTGCAGACAGAGGATGTGTCACCGTGAAGCTTGTAACTC 617
Db 612 CACTGAGGCTGAGAGAGTGCAGACAGGAGATGTGTCAGGATGTGATATG 671
QY 618 CGTTTATTAACACACCTGCGGAGAGATGACAAAGTGAACCAATTGGTATGCTGG 677
Db 672 CGTATGATACACACACATCCGGGAGAGATGACAAAGCTTATGCGGATGCGCG 731
```

```
QY 678 CCTTAATTGAGCGGAGGATCTTGCCTACGCCCGGAGACAGATGATGCTC 737
Db 732 GCCGTTATTGAGCGAGCGGATCTTGCCTATGCCGGAAGACCGAATGATCCATC 791
QY 738 GAAAGCTGC 746
Db 792 CAAGTCTGC 800

RESULT 4
US-09-173-300-50
; Sequence 50, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173.300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 50
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Glycine max
US-09-173-300-50

Query Match
Best Local Similarity 68.7%; Score 291.2; DB 4; Length 1107;
Matches 430; Conservative 1; Mismatches 189; Indels 6; Gaps 2;

112 CTCACCTACCTAAATGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171
Db 171 CCCAAGTCTCAAAACCTTCGCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 230
QY 172 GGGGCGAGCAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 231
Db 231 TCGGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 290
QY 232 AATATGACACGACAGATATCCCGCGAGACCTCACTGAGGAGCTGAGGAGCG 291
Db 291 AATATGACACGACAGATATCCCGCGAGATCAAAACCAAGTACGCACTGCTGCTG 350
QY 292 GACGAGTACCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 351
Db 351 GACGAGTACGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 407
QY 352 AGCGCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 411
Db 408 AGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467
QY 412 AACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 471
Db 468 AACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 527
QY 472 CGCGCATATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531
Db 528 CGCGCATATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 587
QY 532 GAGGTGTACCTCTGAGAGTCAAGGAGTGGGCTGGAAGAGTGAAGAGAGGAT 591
Db 588 GAGGTGTATCCCTGAGAGTTCGAGAGGAGAGCTGCTGCGAGGAGTCAACAGCGGAT 644
QY 592 GTGGTACAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651
Db 645 GTGGTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 704
```



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,420
FILING DATE: December 3, 1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1472 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1105
US-08-781-420-10

Query Match
Best Local Similarity 5.5%; Score 57; DB 4; Length 1472;
Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 913 TTTTGTGTTGCTACAGTACTTGTCATTTGTAATGTCCTGCCTGCTACTGTCTAATCT 972
Db 1351 TTTTGGTTCCCTAGTAGTACTGTGTATTTGTATTAATAATTTTGCCCTGTTTTAAAT 1410
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 973 TTGAATACTGCTCGTGGCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
Db 1411 TTTTAAATAATTTTTTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1470
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1033 A 1033
Db 1471 A 1471

RESULT 9
US-08-781-420-12/C
Sequence 12, Application US/08781420
Patent No. 6248872
GENERAL INFORMATION:
APPLICANT: Chandrasekar, Ramaswamy
APPLICANT: Mehta, Kapil
TITLE OF INVENTION: Parasitic Nematode Transglutaminase
TITLE OF INVENTION: Proteins, Nucleic Acid Molecules and
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,420
FILING DATE: December 3, 1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington

```

```

? ; FEATURE: CDS
? ; NAME/KEY :
? ; LOCATION : 2..1105
US-08-874-102-10

Query Match          5.5%: Score 57; DB 4; Length 1472;
Best Local Similarity 66.9%; Pred No. 0.00035;
Matches      81; Conservative    0; Mismatches   40; Indels     0; Gaps     0;

OY      913 TTTTGTGGTAGTACTGTTGACAAATTGTACTCCTGCCTGCTACTGTTCTTAATCTGT 972
        ||||| ||| - | ||| - ||||| - || - || - ||| - |
DB      1351 TTTTGTTTACTTACTTATAGTACTGTTGTAATGTAATTACTAATTTTGCCTGTGTTTTTAAT 1410
OY      973 TTGAATAACTGCTGCTGTCGCCAAAAAAAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
DB      1411 TTTAAATAAATTTTTTTTGGAAAAAAAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1470
OY      1033 A 1033
DB      1471 A 1471

RESULT 11
US-08-874-102-12/C
; Sequence 12, Application US/08874102
; Patent No. 6309644
GENERAL INFORMATION:
APPLICANT: Ramaswamy Chandrashekar
APPLICANT: Kapil Mehta
TITLE OF INVENTION: PARASITIC NEMATODE TRANSGLOUTRININASE
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES,
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874.102
FILING DATE: 12-JUN-1997
CLASSIFICATION: A24
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: HW-2-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/464-9505
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1472 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-874-102-12

Query Match          5.5%; Score 57; DB 4; Length 1472;
Best Local Similarity 66.9%; Pred. No. 0.00035;
Matches      81; Conservative    0; Mismatches   40; Indels     0; Gaps     0;

OY      913 TTTTGTGGTAGTACTGTTGACAAATTGTACTCCTGCCTGCTACTGTTCTTAATCTGT 972
        ||||| ||| - | ||| - ||||| - || - || - ||| - |
DB      122 TTTGTGTTACTTACTGTTGTAATGTAATTACTAATTTTGCCTGTGTTTTTAAT 63

```



```
; Patent No. 6383774
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; TITLE OF INVENTION: PARASITIC NEMATODE TRANSLUTAMINASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,919A
; FILING DATE: 04-DEC-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-2-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1875 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 24..1508
; US-08-984-919A-46

Query Match          5.5%; Score 57; DB 4; Length 1875;
Best Local Similarity 66.9%; Pred. No. 0.00037;
Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 913 TTTTGTGTTACTAGTACTGTGACAAATTTACTCTGCTGCTACTGTTCTTATCTGT 972
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1754 TTTTGTGTTACTAGTACTGTGTTATTTGTTATTTGCTTGTGTTTAAAT 1813
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 973 TTGAATACGCTGCTGTTGCCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1814 TTTAAATAAATTTTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1873
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1033 A 1033
    |
DB 1874 A 1874

RESULT 15
US-08-984-919A-48/C
; Sequence 48, Application US/08984919A
; Patent No. 6383774
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; TITLE OF INVENTION: PARASITIC NEMATODE TRANSLUTAMINASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heeka Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Wordperfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,919A
; FILING DATE: 04-DEC-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-2-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1875 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-984-919A-48

Query Match          5.5%; Score 57; DB 4; Length 1875;
Best Local Similarity 66.9%; Pred. No. 0.00037;
Matches 81; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 913 TTTTGTGTTACTAGTACTGTGACAAATTTACTCTGCTGCTACTGTTCTTATCTGT 972
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 122 TTTTGTGTTACTAGTACTGTGTTATTTGTTATTTGCTTGTGTTTAAAT 63
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 973 TTGAATACGCTGCTGTTGCCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1032
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 TTTAAATAAATTTTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1033 A 1033
    |
DB 2 A 2

Search completed: April 3, 2003, 04:39:51
Job time : 1467 secs
```

GenCore version 5.1.4.P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:04:48 : Search time 36 Seconds
(without alignments)
921.650 Million cell updates/sec

Title: US-10-027-450-47

Perfect score: 1266

Sequence: 1 MAALSGTAVSTAALAPIR.....AGGIFAYARKTGMIASKAA 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq-101002:*

1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	100.0	249	21	AA198439
2	1226	96.8	244	21	AA198442
3	994.5	78.6	257	21	AA198440
4	740.5	58.5	263	21	AA198441
5	710.5	56.1	251	21	AA198443
6	710.5	56.1	251	21	AA198444
7	710.5	56.1	251	21	AA198445
8	710.5	56.1	251	21	AA198446
9	710.5	56.1	251	21	AA198447
10	651	51.4	253	23	AA198448

11	644.5	50.9	256	23	AB192089
12	627	49.5	256	23	AB192090
13	627	49.5	256	23	AB192091
14	258.5	20.4	164	22	AA196132
15	206.5	16.0	163	21	AA195395
16	202	15.1	191	15	AA194217
17	191.5	15.1	191	15	AA194217
18	191.5	15.1	191	15	AA194217
19	184	14.5	212	22	AA196307
20	174.5	13.8	193	23	AB194173
21	166	13.1	189	23	AB195041
22	164.5	13.0	644	22	AA194234
23	162	12.8	164	18	AA192821
24	155.5	12.3	201	22	AA198423
25	155	12.2	190	22	AA196564
26	153.5	12.1	201	22	AA194427
27	151	11.9	200	22	AA195572
28	134	10.6	197	22	AA191199
29	134	10.6	197	22	AA191199
30	130	10.3	943	22	AA191445
31	127.5	10.1	902	22	AB191080
32	126	10.0	639	21	AA194440
33	126	10.0	868	21	AA193301
34	125.5	9.9	553	22	AA195520
35	125.5	9.9	557	22	AA195519
36	125.5	9.9	639	21	AA194439
37	125.5	9.9	869	21	AA193285
38	125	9.9	639	21	AA194438
39	120	9.5	200	22	AA196055
40	119.5	9.4	139	23	AB192235
41	115.5	9.1	780	20	AA199447
42	115.5	9.1	160	22	AA194333
43	114.5	9.0	160	22	AA194333
44	112	8.8	417	21	AA198362
45	112	8.8	889	23	AB191134

ALIGNMENTS

RESULT 1
AA198439
ID AAY28439 standard; Protein: 249 AA.
XX
AC AAY28439;
XX
DT 15-FEB-2000 (first entry)
XX
DE Corn leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
XX
KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
KW 3-isopropylmalate dehydratase.
XX
OS Zea mays.
XX
PN WO921880-A2.
XX
PD 06-MAY-1999.
XX
PF 20-OCT-1998; 98WO-US22081.
XX
PR 28-OCT-1997; 97US-0063423.
XX
(DUPO) DU POINT DE MEMOIRS & CO E I.
XX
PI Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
DR WPI: 2000-022904/02.
XX N-PSDB: AAX89462.
XX
PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
enzymes -

XX Claim 16; Fig 7; 102pp; English.
 PS
 CC AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 XX
 SQ Sequence 249 AA;
 Query Match 100.0%; Score 1266; DB 21; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.5e-119;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAALSGTAVSTALAPRAPSFIIRSQLTCHRLSLKCRAGSIVPAAAAAGSS 60
 DB 1 MAAALSGTAVSTALAPRAPSFIIRSQLTCHRLSLKCRAGSIVPAAAAAGSS 60
 QY 61 SPSSAVFHGECEFYVDNIDTDQIIPAEHLTLVPSKPEYRKLGSFAFAGLPSSAAYPPFV 120
 DB 61 SPSSAVFHGECEFYVDNIDTDQIIPAEHLTLVPSKPEYRKLGSFAFAGLPSSAAYPPFV 120
 QY 121 AGESSSRATIVGANGCGSSREHAPVALGAGARIVAGYARIFFRNSVATGCVYP 180
 DB 121 AGESSSRATIVGANGCGSSREHAPVALGAGARIVAGYARIFFRNSVATGCVYP 180
 QY 181 LELTVGAMKECKTGDDVTVDLANSVFINTHSGKEYKLKPIGDAGPIEAGGIFAYARKT 240
 DB 181 LELTVGAMKECKTGDDVTVDLANSVFINTHSGKEYKLKPIGDAGPIEAGGIFAYARKT 240
 QY 241 GMTASKAA 249
 DB 241 GMTASKAA 249
 RESULT 2
 AAY28442
 ID AAY28442 standard; Protein: 244 AA.
 XX
 AC AAY28442;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
 XX
 KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase.
 OS
 OS Triticum aestivum.
 XX
 PN WO9921880-A2.
 XX
 PD 06-MAY-1999.
 XX
 PF 20-OCT-1998; 98WO-US22081.
 XX
 PR 28-OCT-1997; 97US-0063423.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Falco SC, Cahoon RE, Hitz WD, Kinney AJ, Rafalski JA;

DR WPI: 2000-022904/02.
 DR N-PSDB: AAX89465.
 XX
 CC Nucleic acid fragments encoding branched chain amino acid biosynthetic
 CC enzymes
 PS
 CC Claim 16; Fig 7; 102pp; English.
 XX
 CC AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 XX
 SQ Sequence 244 AA;
 Query Match 96.8%; Score 1226; DB 21; Length 244;
 Best Local Similarity 99.6%; Pred. No. 1.6e-115;
 Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 9 AVSTALAPRAPSFIIRSQLTCHRLSLKCRAGSIVPAAAAAGSSSPSSAVPH 68
 DB 4 AVSTALAPRAPSFIIRSQLTCHRLSLKCRAGSIVPAAAAAGSSSPSSAVPH 63
 QY 69 GECEFYVDNIDTDQIIPAEHLTLVPSKPEYRKLGSFAFAGLPSSAAYPPFVAPGESSR 128
 DB 64 GECEFYVDNIDTDQIIPAEHLTLVPSKPEYRKLGSFAFAGLPSSAAYPPFVAPGESSR 123
 QY 129 YAIIVGANGFCGSSREHAPVALGAGARIVAGYARIFFRNSVATGCVYPLELDVGA 188
 DB 124 YAIIVGANGFCGSSREHAPVALGAGARIVAGYARIFFRNSVATGCVYPLELDVGA 183
 QY 189 WKECKTGDDVTVDLANSVFINTHSGKEYKLKPIGDAGPIEAGGIFAYARKGMTASKAA 248
 DB 184 WKECKTGDDVTVDLANSVFINTHSGKEYKLKPIGDAGPIEAGGIFAYARKGMTASKAA 243
 QY 249 A 249
 DB 244 A 244
 RESULT 3
 AAY28440
 ID AAY28440 standard; Protein: 257 AA.
 XX
 AC AAY28440;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Rice leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
 XX
 KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase.
 OS
 OS Oryza sativa.
 XX
 PN WO9921880-A2.
 XX
 PD 06-MAY-1999.
 XX
 PF 20-OCT-1998; 98WO-US22081.
 XX
 PR 28-OCT-1997; 97US-0063423.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX XX Palco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
 FI WPI: 2000-022904/02.
 XX DR N-PSDB: AAX89463.
 XX PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PT enzymes -
 XX PS Claim 16; Fig 7; 102pp; English.
 XX CC AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 XX SQ Sequence 257 AA:
 Query Match 78.6%; Score 994.5; DB 21; Length 257;
 Best Local Similarity 77.0%; Pred. No. 4,2e-92;
 Matches 198; Conservative 18; Mismatches 32; Indels 9; Gaps 2;
 QY 1 MAALSGTAVSTA-----ALLAPRAPSPAFIRRSQLTCHRLSLKC--RRASIVPA 51
 DB 1 MAALAAAPALSLAEAPVAVLAPCPPTSPRTFRRSWVALCRPALCKHNSPLTAAVAA 60
 QY 52 AAAAAAGSSSPSSAVFGECEFYVGDNIPTDQIIPAEHLTVPSKPEDEYKLSFAFAGLP 111
 DB 61 AAAAAAGBSTSAGVFHGECEFYVGDNIPTDQIIPAEHLTVPSKPEDEYKLSFAFAGLP 120
 QY 112 SAAPPPPEVAAGESSRVAIVIGANFGCGSSREHAPVALGAAGARAIVAEGYARIFERN 171
 DB 121 TAAYPPPEVAAGETTRVAVIIGANFGCGSSREHAPVALGAAGARAAVAEGYARIFERN 180
 QY 172 SVATGEYVPLELTDVGAMKECKTGVDVTVDLANSVFINTSGEKYKLPIDGAGVIRAG 231
 DB 181 SVATGEYVPLELTDVGAMKECKTGVDVTVDELNCVAINHTSKQYKLPIDGAGVIRAG 240
 QY 232 GIFAFAKRTGMIAKAA 248
 DB 241 GIFAFAKRTGMIAKSA 257
 RESULT 4
 AAY28441
 ID AAY28441 standard; Protein: 263 AA.
 XX AC AAY28441;
 XX DT 15-FEB-2000 (first entry)
 XX XX Leud subunit of 3-isopropylmalate dehydratase amino acid sequence.
 DE XX
 KW Corn, soybean, wheat, rice, dihydroxyacid dehydratase; leuc, leud,
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
 KW 3-isopropylmalate dehydratase.
 XX OS glycine max.
 XX FT Key Location/Qualifiers
 FT Misc-difference 4

FT FT /label= unknown
 XX XX /note= "encoded by TAG"
 XX PN W09921880-A2.
 XX PD 06-MAY-1999.
 XX PF 20-OCT-1998; 98WO-US22081.
 XX PR 28-OCT-1997; 97US-0063423.
 XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX PI Palco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;
 XX DR WPI: 2000-022904/02.
 XX DR N-PSDB: AAX89464.
 XX PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
 PT enzymes -
 XX PS Claim 16; Fig 7; 102pp; English.
 XX CC AAY28439-Y28442 are fragments of corn, soybean, wheat and rice leud
 CC subunit of 3-isopropylmalate dehydratase, amino acid sequences.
 CC Sequences AAX89442-X89465 are nucleic acid fragments that encode all or a
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
 CC in biosynthesis and utilization of branched-chain amino acids. The
 CC nucleic acid sequences can be used to alter the level of expression of a
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can
 CC also be used to obtain a nucleic acid fragment encoding a branched chain
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
 CC to prepare antibodies for detecting the proteins in situ in cells, or in
 CC vitro in cell extracts.
 XX SQ Sequence 263 AA:
 Query Match 58.5%; Score 740.5; DB 21; Length 263;
 Best Local Similarity 62.8%; Pred. No. 2e-66;
 Matches 157; Conservative 24; Mismatches 52; Indels 17; Gaps 5;
 QY 8 TAVTALLAPRAPTSAPFIRRSQLTCHRL-----HSICKRRAGSIVPAAAAAA 56
 DB 20 TRSSAAATVLP---RNLATKLSLSHSHLDPFLSFPTPKSNPRNRVAVSLQTPRAQS 76
 QY 57 AGSSSPSSAVFGECEFYVGDNIPTDQIIPAEHLTVPSKPEDEYKLSFAFAGIPSAAP 116
 DB 77 AASASP-SASFHGLCYVGDNIPTDQIIPAEHLTVPSKPEDEYKLSYALIGLP-ATVA 134
 QY 117 TPFVAPGEESRVAIVIGANFGCGSSREHAPVALGAAGARAIVAEGYARIFERNVATG 176
 DB 135 TRFIEPGEIKTKYAIIVIGANFGCGSSREHAPVALGAAGAAVAEYARIFERNVATG 194
 QY 177 EYVPLELTDVGAMKECKTGVDVTVDLANSVFINTSGEKYKLPIDGAGVIRAGIFAY 236
 DB 195 EYVPLE-SEGRICECTGDDVTVIELGSRULNHTGKRYKLPIDGAGVIRAGIFAY 253
 QY 237 ARKTMIAK 246
 DB 254 ARKTMIPSR 263
 RESULT 5
 AAG26403
 ID AAG26403 standard; Protein: 251 AA.
 XX AC AAG26403;
 XX XX
 XX DT 17-OCT-2000 (first entry)
 XX FT

DE Arabidopsis thaliana protein fragment SEQ ID NO: 30845.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139765.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 02-JUL-1999; 99US-0141287.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145102.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145228.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.


```
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 56.1%; Score 710.5; DB 21; Length 251;
Best Local Similarity 59.5%; Pred. No. 2,1e-63;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

```
QY 1 MAALSGTAVSTALAPTRAPTSFIRRSQLTCHRLSLKCRKRRAGSTVPAAPAAAAGSS 60
DB 1 MAALSGTAVSTALAPTRAPTSFIRRSQLTCHRLSLKCRKRRAGSTVPAAPAAAAGSS 60
QY 61 SPESA-----VFHGECEVVDNDITDQIIIPAEHLTVLPSPDEYRRLGSAFAPLPSA 113
DB 57 VFSRAEPOERKTFHGLICVVDNDITDQIIIPAEHLTVLPSPDEYRRLGSAFAPLPSA 115
QY 114 AYTTPVAFGESESSRAIIVGANGFCGSSREHAPVALGAGARIVAGEYARIFERNRSV 173
DB 116 SYKERVVOGEMKTKSIIITIGENCGSSREHAPVCLGAGAKAVVAGSTARIFERNRSV 175
QY 174 ATGEVTPLELTVYGAMKECKTGDVYTVDI--ANSVFINHTSGKEYKLKPIGDAGPVIAG 231
DB 176 ATGEVTPLELTVYGAMKECKTGDVYTVDI--ANSVFINHTSGKEYKLKPIGDAGPVIAG 234
QY 232 GIFAFAKTMATSKAA 248
DB 235 GIFAFAKTMATSKAA 251
```

```
RESULT 6
AAG42502
ID AAG42502 standard; Protein; 251 AA.
XX
AC AAG42502;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 53014.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0133046.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
```


QY	174	ATGEEVPELEILDYGMAREKRCIKGDPVTVDT--AASVFINHTSGREKIKLPIDGADAPVLEAG	231
	175	: : : : : : : : : : : : : :	
DB	176	ATGEVPLD-SEVRACDECTGGVAIVELKEDSILINHTGKEIKLPIGDGAPVIDAG	234
QY	232	GIFAVARKTGMTASKAA	248
	233	: : : : : : : : : : : : : :	
	235	GIFAVARKKGMIPSAAA	251

RESULT 7
ABB92091
ID ABB92091 standard; Protein; 251 AA.

PT identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
XX
PS Claim 5; SEQ ID NO 1302; 261bp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (AB030790-AB034016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

[illegible]

```

OY 174 ATEBVPLELTIVGAMKECKGDPVYD--ANSVFNHTSGKEKIPISGAPVIEAG 231
OY 175 ATEBVPLELTIVGAMKECKGDPVYD--ANSVFNHTSGKEKIPISGAPVIEAG 232
Db 176 ATGEVYPLD--SEKRVODECTGDAIVELRGDSILNHTGKEKIPISGAPVIDAG 234
OY 232 GIPAYARKGTMTASRAA 248
OY 233 GIPAYARKGTMTASRAA 251
Db

```

RESULT	8
AAG26402	
ID	AAG26402 standard; Protein; 261 AA
XX	
AC	AAG26402;

PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131448.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 13-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-014484.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147205.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159639.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 56.1%; Score 710.5; DB 21; Length 261;
Best Local Similarity 59.5%; Pred. No. 2,2e-63;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

QY 1 MAALSTANVTALAPIRAPTSATIRRSQLTCHRLHSIKCRAGSIVPAAAAAGSS 60
Db 11 MAASLOSANPTLSRTLASPNKPKSSFATFRSPFL--RFNSTSV--ASNFPPLYSREASSSF 66

PR	04-AUG-1999;	9905-014730.4
PR	04-AUG-1999;	9905-014730.2
PR	05-AUG-1999;	9905-014719.2
PR	05-AUG-1999;	9905-014778.0
PR	06-AUG-1999;	9905-014773.0
PR	06-AUG-1999;	9905-014747.6
PR	09-AUG-1999;	9905-014749.3
PR	09-AUG-1999;	9905-014793.5
PR	09-AUG-1999;	9905-014793.1
PR	10-AUG-1999;	9905-014817.5
PR	11-AUG-1999;	9905-014831.9
PR	12-AUG-1999;	9905-014834.1
PR	13-AUG-1999;	9905-014835.5
PR	16-AUG-1999;	9905-014848.8
PR	16-AUG-1999;	9905-014936.8
PR	17-AUG-1999;	9905-014917.5
PR	17-AUG-1999;	9905-014942.6
PR	20-AUG-1999;	9905-014972.2
PR	20-AUG-1999;	9905-014972.3
PR	20-AUG-1999;	9905-014992.9
PR	23-AUG-1999;	9905-014993.0
PR	23-AUG-1999;	9905-014993.0
PR	25-AUG-1999;	9905-015056.6
PR	26-AUG-1999;	9905-015088.4
PR	27-AUG-1999;	9905-015106.5
PR	27-AUG-1999;	9905-015106.6
PR	27-AUG-1999;	9905-015108.0
PR	30-AUG-1999;	9905-015130.3
PR	31-AUG-1999;	9905-015143.8
PR	01-SEP-1999;	9905-015193.0
PR	07-SEP-1999;	9905-015236.3
PR	10-SEP-1999;	9905-015307.0
PR	13-SEP-1999;	9905-015378.8
PR	15-SEP-1999;	9905-015401.8
PR	16-SEP-1999;	9905-015403.9
PR	20-SEP-1999;	9905-015477.9
PR	22-SEP-1999;	9905-015513.9
PR	23-SEP-1999;	9905-015548.6
PR	24-SEP-1999;	9905-015565.9
PR	28-SEP-1999;	9905-015638.6
PR	29-SEP-1999;	9905-015659.6
PR	04-OCT-1999;	9905-015711.7
PR	05-OCT-1999;	9905-015775.3
PR	06-OCT-1999;	9905-015786.5
PR	07-OCT-1999;	9905-015802.9
PR	08-OCT-1999;	9905-015823.2
PR	12-OCT-1999;	9905-015836.6
PR	13-OCT-1999;	9905-015929.3
PR	13-OCT-1999;	9905-015929.4
PR	13-OCT-1999;	9905-015929.5
PR	14-OCT-1999;	9905-015933.1
PR	14-OCT-1999;	9905-015933.1
PR	14-OCT-1999;	9905-015963.7
PR	18-OCT-1999;	9905-015958.4
PR	21-OCT-1999;	9905-016074.1
PR	21-OCT-1999;	9905-016076.7
PR	21-OCT-1999;	9905-016076.8
PR	21-OCT-1999;	9905-016077.0
PR	21-OCT-1999;	9905-016081.4
PR	21-OCT-1999;	9905-016081.5
PR	22-OCT-1999;	9905-016098.0
PR	22-OCT-1999;	9905-016098.1
PR	22-OCT-1999;	9905-016098.9
PR	25-OCT-1999;	9905-016140.4
PR	25-OCT-1999;	9905-016140.5
PR	26-OCT-1999;	9905-016135.9
PR	26-OCT-1999;	9905-016136.0
PR	26-OCT-1999;	9905-016136.1
PR	28-OCT-1999;	9905-016192.0
PR	28-OCT-1999;	9905-016199.3

[illegible]

Query Match 51.4%; Score 651; DB 23; Length 253;
 Best Local Similarity 59.7%; Pred. NO. 2.2e-57;
 Matches 135; Conservative 29; Mismatches 56; Indels 6; Gaps 4;

QY 24 SAEIR-RSOLTCRLHSLKCRAGSIVPAAAAAGSSSPSSAVFHEGCFVYGNIDTDQ 82
 Db 29 SPFLQKASATIFENKPCLTSSATITRYAASSSDSGESTIREPHGLCFYLNKDNIDTDQ 88
 QY 83 IITAEHLTVPSKPDYKRLGSFAFAGLPSAAYFFPEFAPEEESRYALITYGANEGCGS 142
 Db 89 IIAEYGLTIPSTPEDRELGSFALNGLPK-FYNERFVVPBEMSKSVIIGDNGFCGS 147
 QY 143 SREHAPVALGAAGARAIVAEYARIFERNVATGEVPELTDVGAMKECKTGVDVYDLD 202
 Db 148 SRHAPVCLGAAGAKAVAEYARIFERNVATGEVPELTDVGAMKECKTGVDVYDLD 206
 QY 203 ---ANSVFIMHNSGKEYKLPKIGDAGPYEAGGIFAYARKTGMTAS 245
 Db 207 KEDGSSLLINHTTRKEYKLPKIGDAGPYIDAGIFAYARKAGMTIPS 252

RESULT 11

ABB92089
 ID ABB92089 standard; Protein; 256 AA.

AC ABB92089;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1300.

KW Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 comprising aligning and comparing nucleic acid or amino acid sequences
 from plant with nucleic acid or amino acid sequences from non-plant
 organisms -

PS Claim 5; SEQ ID NO 1300; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 256 AA;

Query Match 50.9%; Score 644.5; DB 23; Length 256;
 Best Local Similarity 57.9%; Pred. NO. 1e-56;
 Matches 139; Conservative 27; Mismatches 59; Indels 15; Gaps 6;

QY 1 MAALSGTAVSTAALLAPRAPTSAFIRRSOLTCRLHSLKCRAGSIVPAAAAAGSS 60
 Db 1 MAASLOSANPTLSRLTASPNKPSRATFRPPL--RENSIV--ASNPKPLVSEASSSF 56
 QY 61 SSSSA-----VHEGECFVVGNDNIDTDQIIPAEHLTVPSKPDYKRLGSFAFAGLPSA 113
 Db 57 VTRSAAEPOERKTFPHGLCYVVGNDNIDTDQIIPAEFLTVPSNPEEYKLSYALVGLP-A 115
 QY 114 AYPTEPVAPGEESRYALITYGANEGCGSSREHAPVALGAAGARAIVAEYARIFERNV 173
 Db 116 STKEFVOPGEKTKYISIIIGENEGCGSSREHAPVCLGAAGAKAVAAQSTARIFERNV 175
 QY 174 AGEVYVPELTDVGAMKECKTGVDVYDLD--ANSVFIMHNSGKEYKLPKIGDAGPYEAG 231
 Db 176 AGEVYVPLD-SEVRVDCBECTGDVATVRELREGDSITLINHTTGKEYKLPKIGDAGPYIDAG 234

RESULT 12

ABB92090
 ID ABB92090 standard; Protein; 256 AA.

AC ABB92090;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1301.

KW Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 comprising aligning and comparing nucleic acid or amino acid sequences
 from plant with nucleic acid or amino acid sequences from non-plant
 organisms -

PS Claim 5; SEQ ID NO 1301; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 256 AA;

Query Match 49.5%; Score 627; DB 23; Length 256;
 Best Local Similarity 54.2%; Pred. NO. 5.9e-53;
 Matches 136; Conservative 38; Mismatches 65; Indels 12; Gaps 6;

QY 4 ALSGTAVSTAALLAPRAPTSAFIRRSOLTCRLHSLKCRAG-----SIYPAANAAGS 59
 Db 11 ALPCSSTKSSSLATRSP---FLRFNGSTSLSPSSISITRGRGSSPTIIPRAAASDS 67
 QY 60 SSP-SSAVHFHEGCFVVGNDNIDTDQIIPAEHLTVPSKPDYKRLGSFAFAGLPSAAYPTP 118

Db	68	NEALANTFHELQCYLVKNDIDTDQIIIPGAACTFPSSMOERDEIAAALSLGLD-FHKTR	126
Qy	119	FVAPGESSRAIIIVGANGFCGSSREHAPALGAAARAIVAGYARIFRNSVATGEV	178
Db	127	FIEPENSRSKSIITIGGENFCGSSREHAPVCLGAAAKAIIVASYARIFRNSVATGEV	186
Qy	179	YPLELIDVGAKECKTGQVYTVDDLANS--VFINTSGSEKYLKEPIDGAPVIEAGGIFAY	236
Db	187	PPLP-SEKRVACECKTGDPTVITIELSDSGLLTNHTTGRNKKLSIDGAGVIDAGGIFAY	245
Qy	237	ARKTGMIAASKA	247
Db	246	ARMGMIPSLA	256

RESULT 13

ABB92092
ID ABB92092 standard; Protein; 256 AA.

aa Identifying plant target proteins for herbicidally active compounds,
 pt comprising aligning and comparing nucleic acid or amino acid sequences
 pt from plant with nucleic acid or amino acid sequences from non-plant
 pt organisms -

The invention relates to identifying target proteins, (AB930790-AB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides.

SQ Sequence 256 AA;

Query Match	49.5%	Score 627;	DB 23;	Length 256;
Best Local Similarity	54.2%	Pred. No. 5.9e-55;		
Matches 136; Conservative	38;	Mismatches 65;	Indels 12;	Gaps 6

[illegible]

RESULT 14

AAB96359
ID AAB96359 standard; Protein; 164 AA.

Accession	Protein	Accession	Protein
AA	New nucleotide sequences isolated from <i>Pyrococcus abyssi</i> encode	AA	New nucleotide sequences isolated from <i>Pyrococcus abyssi</i> encode
PT	proteins useful in industry -	PT	proteins useful in industry -

AA The present invention relates to the genomic sequence of Pyrococcus
CC abyssalis (see AAB86431 and AAH1223-7) and P. abyssalis proteins. P. abyssalis is
CC abyssalis (see AAB86431 and AAH1223-7) and P. abyssalis proteins. P. abyssalis is
CC a hyperthermophilic archaeon, which is isolated from deep sea
CC hydrothermal vents. The present sequence is one such P. abyssalis protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAB75903-AAH75920 and AAG66436.

SQ Sequence 164 AA;

Query Match	20.4%	Score 258.5	DB 22	Length 164
Best Local Similarity	39.08	Pred. No. 5.5e-18		
Matches 67; Conservative	26	Mismatches 62	Indels 17	Gaps 7

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:07:03 ; Search time 15 Seconds
(without alignments)
488.421 Million cell updates/sec

Title: US-10-027-450-47

Perfect score: 1266
Sequence: 1 MAALSGTAVSTAALLAPIR.....AGGIFAYARKTGMIASKAA 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTOTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	100.0	249	4	US-09-173-300-47
2	1226	96.8	244	4	US-09-173-300-53
3	994.5	78.6	257	4	US-09-173-300-49
4	740.5	58.5	263	4	US-09-173-300-51
5	191.5	15.1	191	1	US-08-403-866-4
6	166	13.1	189	4	US-09-134-001C-3886
7	115.5	9.1	189	2	US-08-887-798-2
8	98.5	7.8	914	4	US-09-134-001C-5208
9	88	7.0	461	4	US-09-355-115-7
10	87	6.9	119	4	US-08-858-207A-388
11	85.5	6.8	4550	2	US-08-804-227C-8
12	85.5	6.8	4550	2	US-08-804-198-2
13	83	6.6	742	4	US-08-791-115B-5
14	81	6.4	1544	4	US-09-413-814-46
15	80.5	6.4	1234	4	US-08-317-310A-15
16	80.5	6.4	1334	5	PCF-US95-130A1-15
17	79.5	6.3	378	4	US-09-325-932A-158
18	79.5	6.3	3739	3	US-09-320-878-2
19	79	6.2	1864	2	US-08-804-227C-3
20	78	6.2	514	4	US-09-385-028-5
21	78	6.2	622	4	US-09-134-001C-3882
22	78	6.2	731	2	US-08-911-364-1
23	78	6.2	733	4	US-08-464-700-2
24	78	6.2	792	4	US-08-678-039A-40
25	78	6.2	1890	4	US-09-004-838-88
26	77.5	6.1	1243	2	US-08-557-139-2
27	77.5	6.1	4472	2	US-08-804-227C-2

28	77.5	6.1	4545	2	US-08-804-227C-14	Sequence 14, Appl
29	76.5	6.0	430	2	US-08-809-740A-2	Sequence 2, Appl
30	76	6.0	387	4	US-09-323-427-4	Sequence 4, Appl
31	76	6.0	387	4	US-09-812-642-4	Sequence 4, Appl
32	76	6.0	970	1	US-08-375-709-7	Sequence 7, Appl
33	76	6.0	970	1	US-08-752-929-7	Sequence 7, Appl
34	76	6.0	970	4	US-09-090-793-5	Sequence 5, Appl
35	76	6.0	1261	1	US-08-764-100-26	Sequence 26, Appl
36	76	6.0	1854	4	US-09-004-838-108	Sequence 108, App
37	75.5	6.0	591	3	US-08-965-903B-2	Sequence 2, Appl
38	75.5	6.0	1298	2	US-08-690-473-2	Sequence 2, Appl
39	75.5	6.0	1298	4	US-09-259-821A-2	Sequence 2, Appl
40	75.5	6.0	1298	4	US-08-843-659-2	Sequence 2, Appl
41	75	5.9	505	1	US-08-221-750A-5	Sequence 5, Appl
42	75	5.9	3165	2	US-08-459-146-3	Sequence 3, Appl
43	75	5.9	3165	2	US-08-459-065-3	Sequence 3, Appl
44	75	5.9	3739	4	US-09-105-537-33	Sequence 33, Appl
45	75	5.9	11877	4	US-09-105-537-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-09-173-300-47
; Sequence 47, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: Bb-1126
; CURRENT APPLICATION NUMBER: US/09/173, 300
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063, 423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 47
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Zea mays
US-09-173-300-47

Query Match      100.0%  Score 1266; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 4.3e-138;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAALSGTAVSTAALLAPIRAPTSAFIRRSQITCRHLSTLCRRAGSTVPAAGAAAAAGSS 60
    |||||||
DB 1 MAALSGTAVSTAALLAPIRAPTSAFIRRSQITCRHLSTLCRRAGSTVPAAGAAAAAGSS 60

QY 1 SPSSANFHECPYVGNIDTDOITFAEHRTIVPSKPDYRKLGSFAFGLPSAAYPPPV 120
    |||||||
DB 61 SPSSANFHECPYVGNIDTDOITFAEHRTIVPSKPDYRKLGSFAFGLPSAAYPPPV 120

QY 121 APGEESRYAIIIGVANGFCGSSREHAPVALGACARAIVAEGARIFFRNSVATGEVYP 180
    |||||||
DB 121 APGEESRYAIIIGVANGFCGSSREHAPVALGACARAIVAEGARIFFRNSVATGEVYP 180

QY 181 LETTDVGAMKECTGTVVVDLANSVFVNIHNSGKRYKLPICGDGPFVTEAGGIFAARKT 240
    |||||||
DB 181 LETTDVGAMKECTGTVVVDLANSVFVNIHNSGKRYKLPICGDGPFVTEAGGIFAARKT 240

QY 241 GMTASKAAA 249
    |||||||
DB 241 GMTASKAAA 249

RESULT 2
```

```

US-09-173-300-53
; Sequence 53, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Ratajski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; CURRENT FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 53
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-173-300-53

```

Query Match	96.8%	Score 1226;	DB 4;	Length 244;
Best Local Similarity	99.6%	Pred. No. 1.8e-133;		
Matches 240; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	9	AVSTALLAPIRAPPIPSAIFRRSOLTCRHLHSLKCRAGSIYVAAAAAGSSSPSSAVNH	68
Db	4	AVSTALLAPIRAPPIPSAIFRRSOLTCRHLHSLKCRAGSIYVAAAAAGSSSPSSAVNH	63
QY	69	GCFCVVDNIDIDDOIIPAEHLTLVPSKPDPEYKLGSEFAGLPSAYPFVPAFGESSR	128
Db	64	GCFCVVDNIDIDDOIIPAEHLTLVPSKPDPEYKRLGSEFAGLPSAYPFVPAFGESSR	123
QY	129	VAIIVGGANFGCGSSREHAPVALGAAGARAIVAEGTARIFRRSVATGVTPELTDVGA	188
Db	124	VAIIVGGANFGCGSSREHAPVALGAAGARAIVAEGTARIFRRSVATGVTPELTDVGA	183
QY	189	WKECKTGDVYTVDLANSVEFINHTSGKEYKLKPIGDAGPYIEAGGIFAAYARKTGMISKAA	248
Db	184	WKECKTGDVYTVDLANSVEFINHTSGKEYKLKPIGDAGPYIEAGGIFAAYARKTGMISKAA	243
QY	249	A 249	
Db	244	A 244	

RESULT 3
US-09-173-300-49
; Sequence 49, Application US/09173300
; Patent No. 6451581

```

? APPLICANT: Falco, Saverio Carl
? APPLICANT: Hitz, William D.
? APPLICANT: Kinney, Anthony J.
? APPLICANT: Cahoon, Rebecca E.
? APPLICANT: Rafalski, J Antoni
? TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC
? ENZYMS
? FILE REFERENCE: BB-1126
? CURRENT APPLICATION NUMBER: US/09/173,300
? CURRENT FILING DATE: 1998-10-15
? EARLIER APPLICATION NUMBER: 60/063,423
? EARLIER FILING DATE: 1997 October 28
? NUMBER OF SEQ ID NOS: 54
? SOFTWARE: Microsoft Word Version 7.0A
? SEQ ID NO 49
? LENGTH: 257
? TYPE: prt
? ORGANISM: Oryza sativa
? US-09-173-300-49

```

Query Match	78.68; Score 994.5; DB 4; Length 257;
-------------	---------------------------------------

	Best Local Similarity	77.0%:	Pred. No. 1,1e-106;	Matches 199;	Conservative 18;	Mismatches 32;	Indels 9;	Gaps 2;
Qy	1	MAAALSGTAVSTA-----	ALLAPIRAPTSARIFRSQTLCHRLSLKC---RRAGSTVPA	51				
Db	1	MAAATAAAALSTLAEAAPVTAVLAPCPPTSPSTRFRRSRWAAALCRAKALCHHSRPLTAVAA	60					
Qy	52	AAAAAGSSPSSSAVHGECEFYVGNIDPDQILPAEHLLVPSKPDDEKLGSAFAFGP	111					
Db	61	AAAAAAAGDSTSGVGHRECEFYVGNIDPDQILPAEHLLVPSKPDDEKLGSAFAFGP	120					
Qy	112	SAAPTEPFAVPAGESSRYAIIVGAGNFGCGSSREHAPVALGAAGARAIIVAEGARIFERN	171					
Db	121	TAAPIPTFPAVAGEETTRAVIIGGANGCGSSREHAPVALGAAGARAVAEGARIFERN	180					
Qy	172	SVATGEVYPLELIDVCAKMECKTGDDVYVYDLANSFIIHNTSKSEKYLKPIGDAAPIVLEAG	231					
Db	181	SVATGEVYPLELIDTGMKMECKTGDDVYVELDNCVMINHSTGKQYKLPIDGAAPIVLEAG	240					
Qy	232	GIFAYARKTGMIASKAA	248					
Db	241	GIFAYARKTGMIASKSA	257					

RESULT 4
US-09-173-300-51
; Sequence 51, Application US/09173300

```

/ Patent No. 6451581
/ GENERAL INFORMATION:
/ APPLICANT: Falco, Saverio Carl
/ APPLICANT: Hitez, William D.
/ APPLICANT: Kinney, Anthony J.
/ APPLICANT: Cahoon, Rebecca E.
/ APPLICANT: Rafalski, J. Antoni
/ TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
/ FILE REFERENCE: BB-1126
/ CURRENT APPLICATION NUMBER: US/09/173,300
/ CURRENT FILING DATE: 1998-10-15
/ EARLIER APPLICATION NUMBER: 60/063,423
/ EARLIER FILING DATE: 1997 October 28
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: Microsoft Word Version 7.0A
/ SEQ ID NO 51
/      LENGTH: 263
/      TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: UNSURE
/      LOCATION: (4)
/ US-09-173-300-51

```

Query Match	58.58;	Score 740.5;	DB 4;	Length 263;
Best Similarity	62.88;	Pred. No. 2.5e-77;		
Best Local	21;	Mismatches 52;	Indels 17;	Gaps 5;
Matches 157;	Conservative			

```

QY      8 TAASTALLIAPRAPPTSAFIRRSQTCRL-----HSLKCRASSIYPMAAAAA 56
      11 : : : : :
Db     20 TRRSSAATLP--RNLAFTKLISTSHSTLTPRELSPETPKSSPNRRNVAVSLOTPAQS 76
      11 : : : : :
QY     57 AGSSSPSAAVFHGCEFFVAGNDINTDOIIPAEHTLVPSKPDYERKLGSFAPAGLSPAYP 116
      11 : : : : :
Db     77 AASASP-SASFHGLCYVAGNDINTDOIIPAEHTLVPSKPDYERKLSTALIGLP-ATYA 134
      11 : : : : :
QY    117 TTPVAPGESSRAIIVGANGFCSSSREHAPVALGAGARAIYAEGYARFFENSVATG 176
      11 : : : : :
Db    135 TRIEGELIKTKAIYIGSANGCGSSREHAPVALGAGAAVAEASARYFFERNVATG 194
      11 : : : : :
QY    177 EYVPELTDVGAMKEKTDVYTVLANSVFNHTSGEKYLLKPGDAPGYIEAGGIFAY 236
      11 : : : : :
Db    195 EYVPLE-SEGRICECCTGDVVTIELGESRLNHTTGKRYLKLPGDAGPYIEAGGIFAY 255
      11 : : : : :
QY    237 ARRTGMIASK 246
      11 : : : : :

```

Db 254 ARKGMIPSR 263

RESULT 5

US-08-403-866-4

; Sequence 4, Application US/08403866

; Patent No. 5643779

; GENERAL INFORMATION:

; APPLICANT: Ehrlich, Stanislaw

; APPLICANT: Godon, Jean-Jacques

; APPLICANT: Renault, Pierre

; TITLE OF INVENTION: Nucleic acid coding for an alpha-acetolactate

; TITLE OF INVENTION: synthase from Lactococcus and its applications

; NUMBER OF SEQUENCES: 16

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/403,866

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30, 727

; REFERENCE/DOCKET NUMBER: 20747/30

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1487

; TELEFAX: (716) 263-1487

; TELEX: 978450 (WUT)

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 191 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Lactococcus lactis subsp. lactis

; INDIVIDUAL ISOLATE: LEUD

; US-08-403-866-4

Query Match

Best Local Similarity 15.1%; Score 191.5; DB 1; Length 191;

Matches 58; Conservative 35; Mismatches 70; Indels 37; Gaps 7;

QY 58 GSSPSAVHGEFVGNIDIDIIIPAEHLTVPSKDEYRKLGSFAPLPSAAPT 117

Db 9 GTSTP-----VANDIDIDIIIPAEHLTVPSKDEYRKLGSFAPLPSAAPT 56

QY 118 PVAAGESSRYAIVGANGGSSREHAPVALGAGARAIVAEGYARIFFRNSVANGE 177

Db 57 DFILNARKYKASILLISGDFGSSREHAPVALGAGARAIVAEGYARIFFRNSVANGE 115

QY 178 VYPL-----ELTDVGAKKEKTDGVTVDLANSVFNNHSGKEYKIKPL----- 221

Db 116 LLLPKQREVLMOTKISSOE-----TTIDLPQLITSLGDFHFEIDPIWKDLING 169

QY 222 -GDAGPYIE-AGGIFAYAR 239

Db 170 LDDIGITLQYEAMISAEYK 189

RESULT 6

US-09-134-001C-3886

; Sequence 3886, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3886

; LENGTH: 189

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

; US-09-134-001C-3886

Query Match

Best Local Similarity 13.1%; Score 166; DB 4; Length 189;

Matches 52; Conservative 11; Mismatches 56; Indels 16; Gaps 5;

QY 76 DNIDTQIIPAEHLTVPSKDEYRKLGSFAP---AGLPSAAPTIPVAPGESSRYAIT 132

Db 19 DNIDTQIIPKVLKRVSKS---GFGPFAFDEMRRLPDGSDNDPDKPEYHGASITL 73

QY 133 VGGANGGSSREHAPVALGAGARAIVAEGYARIFFRNSVATGCVPLELTDVGAMKE- 191

Db 74 ITGDNFGCGSSREHAPVALGAGARAIVAEGYARIFFRNSVATGCVPLELTDVGAMKE- 128

QY 192 --CKTGDVTVDLAN 204

Db 129 EHLAQFDEITVDLPN 143

RESULT 7

US-08-887-798-2

; Sequence 2, Application US/08887798

; Patent No. 5922536

; GENERAL INFORMATION:

; APPLICANT: Mayeux, Richard

; APPLICANT: Graziano, Joseph H.

; APPLICANT: Freyer, Greg

; TITLE OF INVENTION: PARKINSON'S DISEASE TESTS

; NUMBER OF SEQUENCES: 38

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/887,798

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P.

; REGISTRATION NUMBER: 28, 678

; REFERENCE/DOCKET NUMBER: 0575/51949/JPW

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-278-0400

; TELEFAX: 212-391-0525

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 780 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-887-798-2

Query Match

Best Local Similarity 9.1%; Score 115.5; DB 2; Length 780;

Matches 23.3%; Pred. No. 0.00022;

[illegible]

```

1 EARLIER FILING DATE: 1997-01-24
2 NUMBER OF SEQ ID NOS: 11
3 SOFTWARE: PatentIn Ver. 2.0
4 SEQ ID NO: 7
5 LENGTH: 461
6 TYPE: PR1
7 ORGANISM: Barley
8 US-09-355-115-7
9
10 Query Match 7.0%; Score 88; DB 4; Length 461;
11 Best Local Similarity 22.6%; Pred. No. 0.15; Indels 80; Gaps 12;
12 Matches 61; Conservative 28; Mismatches 101;
13
14 13 AALLAPRAPTSAFIRRSOLTCHRLHSICKRAGSIYPAAAAAGSSSPSSAVF---HG 69
15 5 AAHNLGRVPPAPFRR-----RARRRRPAAVALATSAAPQRETDKRVITGMG 56
16
17 QY 70 ECFVGNIDT--DQIIPAHLTVPSKPDVEYRKIGSAFAFGLPSSAVYPPFV----- 120
18 Db 57 LASVFGSDVTFYDRLLAGES-----GVCPIDRFDASSFPFRFAGQIRGS 102
19
20 QY 121 APG-----EESRYAIIVG-----GANFGCGSSREHAPVALGAA----- 154
21 Db 103 SEGVIDKNDRLDDCIRCILSGKKALESAGLGAGSD-AHVKLIDVRAGVLVGTGMGL 161
22
23 QY 155 -----GARAIVACGYARI--FFRNSVAVGEVYPLELTVGAM-----KECKGQVTV 200
24 Db 162 SVFSDGVQNLIEGKYRISFPFIPYATIMGSGALLAIDVGMGPNSISTRACATSNICFY 221
25
26 QY 201 DLANSV-----FINHTSGKEYKIKPIGDAG 225
27 Db 222 AAANHIRGEADIIIVAGTFEALIIPIGLCG 251
28
29 RESULT 10
30 US-08-858-207A-388
31 Sequence 388, Application US/08858207A
32 Patent No. 6348328
33
34 GENERAL INFORMATION:
35 APPLICANT: Black, Michael
36 APPLICANT: Hodgson, John
37 APPLICANT: Knowles, David
38 APPLICANT: Nicholas, Richard
39 APPLICANT: Stodola, Robert
40 TITLE OF INVENTION: No. 6348328el Compounds
41 NUMBER OF SEQUENCES: 552
42
43 CORRESPONDENCE ADDRESS:
44 ADDRESSEE: SmtInkline Beecham Corporation
45 STREET: 709 Swedeland Road
46 CITY: King of Prussia
47 STATE: PA
48 COUNTRY: USA
49 ZIP: 19406-0939
50
51 COMPUTER READABLE FORM:
52 MEDIUM TYPE: Diskette
53 OPERATING SYSTEM: DOS
54 SOFTWARE: FastSeq for Windows Version 2.0
55 CURRENT APPLICATION DATA:
56 APPLICATION NUMBER: US/08/858,207A
57 FILING DATE: 09-MAY-1997
58
59 CLASSIFICATION: 435
60 PRIOR APPLICATION DATA:
61 APPLICATION NUMBER: 60/017670
62 FILING DATE: 14-MAY-1996
63 ATTORNEY/AGENT INFORMATION:
64 NAME: Gimm1, Edward R
65 REGISTRATION NUMBER: 38,891
66 REFERENCE/DOCKET NUMBER: P50475
67 TELECOMMUNICATION INFORMATION:
68 TELEPHONE: 610-270-4478
69 TELEFAX: 610-270-5090
70 TELEX:

```

;; INFORMATION FOR SEQ ID NO: 388:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 119 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: NO. 6348328e
US-08-858-207A-388

Query Match
Best Local Similarity 27.9%; Score 87; DB 4; Length 119;
Pred. No. 0.025; Mismatches 50; Indels 14; Gaps 3;

Matches 31; Conservative 16; Mismatches 50; Indels 14; Gaps 3;

QY 141 GSSREHAPVALGAGARAIYAEGRARIFFRNSVATGE---VYPLELTDVGAKCKCTGDPV 197
DB 3 GSSRDHAMALADYGFVYIAGSFGDIHYNNELNGLMLPTVQPREVREKLA--QLKPTDQ 60
QY 198 VTVDLANSVFIN-----HTSGKEYLKPIGDAGPVIAGGIFAYARK 239
DB 61 VTVDLEOQKITISPEEFTFEIDSKRWKHLNLSLDDIGITLYEELIAYEK 111

RESULT 11

US-08-804-227C-8
Sequence 8, Application US/08804227C
Patent No. 5876991

GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuntz, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide

US-08-804-227C-8

Query Match
Best Local Similarity 23.8%; Score 85.5; DB 2; Length 4550;
Pred. No. 9.6; Mismatches 91; Indels 87; Gaps 13;

Matches 64; Conservative 27; Mismatches 91; Indels 87; Gaps 13;

QY 44 RAGSIVPAAAAAGSSPSSAVF---HGECFYVGNIDTDOI-----IPAHLTLVP 93
DB 273 REGGEAVLRACACQARVDPAEVRFEVHGTGTPVDPAEAMALGAVHSGRPADPLVIG 332
QY 94 SKPDEYRKL-GSFAFAGLPSAAV-----PTPRV-----PRLLSAR-- 120

DB 333 SVKTNIGHLEGANGIAGLVKALCLREPTLPGLNFPATPSAIPDQLRLKVOTAAALP 392
QY 121 -ARGEESRAIIVGANGNCGSSR-----EHAP-----VALGAA-----GARAI 159
DB 393 LAPGA---PILAGVSSFGIGGTNCHVYLEHLPSPRTPPAVSASLPDVPPLLSAR-- 446
QY 160 VAEYARIFEFRNSVATGEVYPLELTDVGAKCKCTGDPVTVDLANSVFINTSGKEYLTK 219
DB 447 -SEGALR---AQAVRIGEV-----VERVGDPR-----DVAVSLASTRTLEHRAVY 489
QY 220 PIGDAGPVIAGGIFAYARKTGMIAKAA 248
DB 490 PCGGREGELVALAGLGAAGRVSGVRSRA 518

RESULT 12

US-08-804-198-2
Sequence 2, Application US/08804198
Patent No. 5945320

GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuntz, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLANTENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide

US-08-804-198-2

Query Match
Best Local Similarity 23.8%; Score 85.5; DB 2; Length 4550;
Pred. No. 9.6; Mismatches 91; Indels 87; Gaps 13;

Matches 64; Conservative 27; Mismatches 91; Indels 87; Gaps 13;

QY 44 RAGSIVPAAAAAGSSPSSAVF---HGECFYVGNIDTDOI-----IPAHLTLVP 93
DB 273 REGGEAVLRACACQARVDPAEVRFEVHGTGTPVDPAEAMALGAVHSGRPADPLVIG 332
QY 94 SKPDEYRKL-GSFAFAGLPSAAV-----PTPRV-----PRLLSAR-- 120
DB 333 SVKTNIGHLEGANGIAGLVKALCLREPTLPGLNFPATPSAIPDQLRLKVOTAAALP 392
QY 121 -ARGEESRAIIVGANGNCGSSR-----EHAP-----VALGAA-----GARAI 159
DB 393 LAPGA---PILAGVSSFGIGGTNCHVYLEHLPSPRTPPAVSASLPDVPPLLSAR-- 446

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:08:58 ; Search time 15 Seconds
(without alignments)

974.813 Million cell updates/sec

Title: US-10-027-450-47

Sequence: 1266
1 MAALSGTAVSTALLAPIR.....ACGIFAVARKTGIMASKAA 249

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 segs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	100.0	249	US-10-027-450-47	Sequence 47, Appl
2	1226	96.8	244	US-10-027-450-53	Sequence 53, Appl
3	994.5	78.6	257	US-10-027-450-49	Sequence 49, Appl
4	740.5	58.5	263	US-10-027-450-51	Sequence 51, Appl
5	184	14.5	212	US-09-815-242-11900	Sequence 11900, A
6	164.5	13.0	644	US-09-815-242-5730	Sequence 5730, Ap
7	155.5	12.3	201	US-09-815-242-14016	Sequence 14016, A
8	155	12.2	190	US-09-815-242-12157	Sequence 12157, A
9	153.5	12.1	201	US-09-815-242-10020	Sequence 10020, A
10	151	11.9	200	US-09-815-242-11165	Sequence 11165, A
11	134	10.6	197	US-09-738-626-4953	Sequence 4953, Ap
12	130	10.3	943	US-09-738-626-5199	Sequence 5199, Ap
13	120	9.5	200	US-09-815-242-11648	Sequence 11648, A
14	112	8.8	417	US-09-925-302-700	Sequence 700, Appl
15	105.5	8.3	538	US-09-779-307-2	Sequence 2, Appl
16	105.5	8.3	538	US-09-779-307-12	Sequence 11, Appl
17	101	8.0	547	US-09-779-307-11	Sequence 13, Appl
18	101	8.0	547	US-09-779-307-13	Sequence 12, Appl
19	98	7.7	963	US-09-924-396B-20	Sequence 20, Appl

20	94.5	7.5	1051	9	US-09-976-059-13	Sequence 13, Appl
21	93	7.3	952	9	US-09-924-396B-18	Sequence 18, Appl
22	90	7.1	444	10	US-09-864-761-42633	Sequence 42633, A
23	88.5	7.0	341	9	US-09-738-626-4543	Sequence 4543, Ap
24	88.5	7.0	585	9	US-09-738-626-6032	Sequence 6032, Ap
25	84.5	6.7	324	9	US-09-738-626-4887	Sequence 4887, Ap
26	84	6.6	354	10	US-09-825-414-70	Sequence 70, Appl
27	83	6.6	537	9	US-09-738-626-4284	Sequence 4284, Ap
28	82.5	6.5	1056	9	US-10-161-510-10	Sequence 10, Appl
29	81	6.4	442	10	US-09-749-728B-11	Sequence 11, Appl
30	81	6.4	562	10	US-09-815-242-12159	Sequence 12159, A
31	81	6.4	562	10	US-09-815-242-12929	Sequence 12929, A
32	80.5	6.4	567	9	US-09-344-882-30	Sequence 30, Appl
33	80	6.3	351	10	US-09-815-242-11984	Sequence 11984, A
34	79	6.2	345	10	US-09-789-836-5	Sequence 5, Appl
35	79	6.2	430	9	US-10-029-180-32	Sequence 32, Appl
36	78	6.2	354	10	US-09-909-162-14	Sequence 14, Appl
37	77.5	6.1	398	9	US-09-738-626-6532	Sequence 6532, Ap
38	77.5	6.1	460	9	US-09-976-673-16	Sequence 16, Appl
39	77.5	6.1	470	9	US-09-976-673-18	Sequence 18, Appl
40	77.5	6.1	1242	9	US-09-436-184-5	Sequence 5, Appl
41	77.5	6.1	1242	10	US-09-903-248-5	Sequence 5, Appl
42	77.5	6.1	1242	10	US-09-859-604-5	Sequence 5, Appl
43	77.5	6.1	1242	10	US-09-903-063-5	Sequence 5, Appl
44	77.5	6.1	1242	10	US-09-903-216-5	Sequence 5, Appl
45	77.5	6.1	1242	10	US-09-903-199-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-10-027-450-47
Sequence 47, Application US/10027450
Patent No. US20020102715A1
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hitz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/10/027, 450
PRIOR APPLICATION NUMBER: 2001-12-20
PRIOR FILING DATE: 1997 October 28
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 47
LENGTH: 249
TYPE: PRT
ORGANISM: Zea mays
US-10-027-450-47

Query Match 100.0%; Score 1266; DB 12; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.2e-113;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAALSGTAVSTALLAPIRAPTSAFIRRSQUTCHRLSLKCRAGSIVPAAAAAGSS 60
DB 1 MAALSGTAVSTALLAPIRAPTSAFIRRSQUTCHRLSLKCRAGSIVPAAAAAGSS 60
QY 61 SPSSAVFHECEVVDNDITDQILIPAEHLTVPSKPDYRKLGSFAAGLPSAAYPPPV 120
DB 61 SPSSAVFHECEVVDNDITDQILIPAEHLTVPSKPDYRKLGSFAAGLPSAAYPPPV 120
QY 121 APEESRAIIVGANGCSCSSREHAPVALCAAGARIVAGARIFFRNSVATGYVP 180
DB 121 APEESRAIIVGANGCSCSSREHAPVALCAAGARIVAGARIFFRNSVATGYVP 180
QY 181 LELTDVGAKCKETDQVTVDLANSVFTNHTSGKEYKIKPIGDAGPVIEAGGIFAVARKT 240
DB 181 LELTDVGAKCKETDQVTVDLANSVFTNHTSGKEYKIKPIGDAGPVIEAGGIFAVARKT 240

Db 181 LELFDVGAMKECKTGDVTVDLANSVFINTSGKEYKLKPIGDAGPVIIEAGIFAVARKT 240
 Oy 241 GMIASKAA 249
 Db 241 GMIASKAA 249

RESULT 2
 US-10-027-450-53
 ; Sequence 53, Application US/10027450
 ; Patent No. US20020102715A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Falco, Saverio Carl
 ; APPLICANT: Hiltz, William D.
 ; APPLICANT: Kinney, Anthony J.
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Rafalski, J. Antoni
 ; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
 ; FILE REFERENCE: BB-1126
 ; CURRENT FILING DATE: 2001-12-20
 ; PRIOR APPLICATION NUMBER: 60/063,423
 ; PRIOR FILING DATE: 1997 October 28
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: Microsoft Word Version 7.0A
 ; SEQ ID NO 53
 ; LENGTH: 244
 ; TYPE: PRT
 ; ORGANISM: Triticum aestivum
 US-10-027-450-53

Query Match 96.8%; Score 1226; DB 12; Length 244;
 Best Local Similarity 99.6%; Pred. No. 7.9e-110;
 Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 9 AVSTALLAPIRAPTSAFIRRSQLTCHRLSHLCKRAGSIYPAAAAAGSSSSSAVFH 68
 Db 4 AVSTALLAPIRAPTSAFIRRSQLTCHRLSHLCKRAGSIYPAAAAAGSSSSSAVFH 63
 Oy 69 GECFVGNIDTDIIPAEHLTLVPSKDEYRKLGSAFAGLPSAAYPTFPVAPGESSR 128
 Db 64 GECFVGNIDTDIIPAEHLTLVPSKDEYRKLGSAFAGLPSAAYPTFPVAPGESSR 123
 Oy 129 YAIIVGANFGCGSSREHAPALGAAGARAIYAGYARIFFRNSVATGEVYPLELTVGA 188
 Db 124 YAIIVGANFGCGSSREHAPALGAAGARAIYAGYARIFFRNSVATGEVYPLELTVGA 183
 Oy 189 WKECKTGDVTVDLANSVFINTSGKEYKLKPIGDAGPVIIEAGIFAVARKTGMIASKAA 248
 Db 184 WKECKTGDVTVDLANSVFINTSGKEYKLKPIGDAGPVIIEAGIFAVARKTGMIASKAA 243
 Oy 249 A 249
 Db 244 A 244

RESULT 3
 US-10-027-450-49
 ; Sequence 49, Application US/10027450
 ; Patent No. US20020102715A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Falco, Saverio Carl
 ; APPLICANT: Hiltz, William D.
 ; APPLICANT: Kinney, Anthony J.
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Rafalski, J. Antoni
 ; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
 ; FILE REFERENCE: BB-1126
 ; CURRENT FILING DATE: 2001-12-20
 ; PRIOR APPLICATION NUMBER: 60/063,423
 ; PRIOR FILING DATE: 1997 October 28
 ; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: Microsoft Word Version 7.0A
 ; SEQ ID NO 49
 ; LENGTH: 257
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 US-10-027-450-49

Query Match 78.6%; Score 994.5; DB 12; Length 257;
 Best Local Similarity 77.0%; Pred. No. 1.3e-87;
 Matches 198; Conservative 18; Mismatches 32; Indels 9; Gaps 2;

Oy 1 MAALSGTAVSTA-----ALLAPIRAPTSAFIRRSQLTCHRLSHLCK----RRAGSIYPA 51
 Db 1 MAALSGTAVSTA-----ALLAPIRAPTSAFIRRSQLTCHRLSHLCK----RRAGSIYPA 60
 Oy 52 AAAAAAGSSSSSAVFHGECEVVDNIDTDIIPAEHLTLVPSKDEYRKLGSAFAGLPSA 111
 Db 61 AAAAAAGSSSSSAVFHGECEVVDNIDTDIIPAEHLTLVPSKDEYRKLGSAFAGLPSA 120
 Oy 112 SAAYPTFPVAPGESSRYAIIIVGANFGCGSSREHAPALGAAGARAIYAGYARIFFRN 171
 Db 121 TAAYPTFPVAPGEETTRAYAVIIGANFGCGSSREHAPALGAAGARAIYAGYARIFFRN 180
 Oy 172 SVATGEVYPLELTVGAMKECKTGDVTVDLANSVFINTSGKEYKLKPIGDAGPVIIEAG 231
 Db 181 SVATGEVYPLELTVGAMKECKTGDVTVDLNCVMINTSGKYKLKPIGDAGPVIIEAG 240
 Oy 232 GIFAVARKTGMIASKAA 248
 Db 241 GIFAVARKTGMIASKAA 257

RESULT 4
 US-10-027-450-51
 ; Sequence 51, Application US/10027450
 ; Patent No. US20020102715A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Falco, Saverio Carl
 ; APPLICANT: Hiltz, William D.
 ; APPLICANT: Kinney, Anthony J.
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Rafalski, J. Antoni
 ; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
 ; FILE REFERENCE: BB-1126
 ; CURRENT FILING DATE: 2001-12-20
 ; PRIOR APPLICATION NUMBER: 60/063,423
 ; PRIOR FILING DATE: 1997 October 28
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: Microsoft Word Version 7.0A
 ; SEQ ID NO 51
 ; LENGTH: 263
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (4)
 US-10-027-450-51

Query Match 58.5%; Score 740.5; DB 12; Length 263;
 Best Local Similarity 62.8%; Pred. No. 2.8e-63;
 Matches 157; Conservative 24; Mismatches 52; Indels 17; Gaps 5;
 Oy 8 TAVSTALLAPIRAPTSAFIRRSQLTCHRL-----HSLCKRAGSIYPAAAAAA 56
 Db 20 TRFSSAATVP---RNLATTKISLSHSHLLRFLSFPTPKSSNPRNRVAASLQTPRAOS 76
 Oy 57 AGSSSSSAVFHGECEVVDNIDTDIIPAEHLTLVPSKDEYRKLGSAFAGLPSAAYP 116
 Db 77 AASASP-SASFGDLGVVDNIDTDIIPAEHLTLVPSKDEYRKLGSAFAGLPSAAYP 134
 Oy 117 TFPVAPGESSRYAIIIVGANFGCGSSREHAPALGAAGARAIYAGYARIFFRNSVATG 176

```
Db 135 TREIEPEIKTYAIVIGANGFCSSREHAPVALGASGAAMVAESYARIFERNSVATG 194
Qy 177 EYVPLELTGYAMKECKTGDVYTVDLANSVFINTHSGKEYIKRPGDGPTEIGGTFAY 236
Db 195 EYVPLE-SEBRLCECTGSDVYTVIELGSRLLNHTTGKEYIKRPGDGPVIEAGIIPAY 253
Qy 237 ARKGTMIASK 246
Db 254 ARKGTMIASK 263

RESULT 5
US-09-815-242-11900
; Sequence 11900, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11900
; LENGTH: 212
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11900

Query Match 14.5%; Score 184; DB 10; Length 212;
Best Local Similarity 32.1%; Pred. No. 4.2e-10;
Matches 53; Conservative 23; Mismatches 45; Indels 44; Gaps 7;

Qy 77 NIPDDOITPAHELTLY-----PSKPEYRKIGSFAPGLPAAVPTFVVAAGESS--- 127
Db 18 NVDTDOITPQPLKSIKRGFGPNLFDEKRYL-----DVGOPGDNSKRP 62
Qy 128 -----RY--AIIVGANFGCGSSREHAPVALGAAGARAIVAGYARIFERN SVA 174
Db 63 LNPDEVLPNRYOGASVLLARENFVGCGSSREHAPVALDEYGRFVIADSYADIEFFNNSFK 122
Qy 175 TGEVYPL-----ELTPVGAMKECKTGDVYTVDLANSVFINTHSGK 214
Db 123 NG-LIPITILPEAYDELFRQVANEBSYQSLIDLAQT-VTRPDGK 165

RESULT 6
US-09-815-242-5730
; Sequence 5730, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5730
; LENGTH: 644
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(644)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5730

Query Match 13.0%; Score 164.5; DB 10; Length 644;
Best Local Similarity 33.3%; Pred. No. 1.4e-07;
Matches 53; Conservative 12; Mismatches 51; Indels 43; Gaps 5;

Qy 41 KCRRAGSIVPAAAAAASSSSPSAVFNGCEV-----VG 75
Db 427 KCARHTLVSPAMAAAA-----IHGK-FDVRRKVVVXMAIKPITTYKGIYPLFN 476
Qy 76 DNIDDOITPAHELTLYPSKPEYRKIGSFAP---AGLPAAVPTFVVAAGESSRYATI 132
Db 477 DNIDDOITPQPLKSIKRGFGPNLFDEKRYL-----DVGOPGDNSKRP 62
Qy 133 VGANFGCGSSREHAPVALGAAGARAIVAGYARIFERN 171
Db 532 ITGDNFGCGSSREHAPVALKDYGFHIIYAGSFSDIFYWN 570

RESULT 7
US-09-815-242-14016
; Sequence 14016, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
```


Db 104 PSFADIFYGNSF--NNGLFVKLSD 126

RESULT 10
US-09-815-242-11165
; Sequence 1165, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11165
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11165

Query Match 11.9%; Score 151; DB 10; Length 200;
Best Local Similarity 28.2%; Pred. No. 5,6e-07;
Matches 48; Conservative 21; Mismatches 59; Indels 42; Gaps 6;

QY 45 AGSIVRAAAAAAGSSPSAFAHGECPYVGNITDQIIIAEHLTLVPSKP-----DE 98
DB 8 SGLVPLDAA-----NVDPAITPKQFLQATIRIGFGKHLFHE 45
QY 99 YRLKGSFAPAGLPSAAYPTP-FVAPGESSRYAIIIVGANFGCGSSRHHAPVALGAGAR 157
DB 46 WRYLD-----VEGTKPMEDEVLNPOYGATFILLARKNIGCGSSRHHAPVALADYGFK 98
QY 158 AIIAEGYARIFRNSVATGEVYPL-----ELPDVGAKMECKTGVDVYVDL 202
DB 99 VMTAPSFADIFYNNSL--NNHMLPIRLSEBEVEITQWVWANEGRQIHWDL 147

RESULT 11
US-09-738-626-4953
; Sequence 4953, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO

; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4953
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4953

Query Match 10.6%; Score 134; DB 9; Length 197;
Best Local Similarity 30.2%; Pred. No. 2,3e-05;
Matches 52; Conservative 20; Mismatches 62; Indels 38; Gaps 6;

QY 77 NIDTDQIIIAEHLTLVPSKPDEYRKLGSFAPAGLPSAAYPT-----PVAAGESSRYA 130
DB 18 MNDTDQIIIAVYLVKRYTR-----TGFEDGLFSWMRONDPFVLMDDTYKNGS 64
QY 131 IIVGANFGCGSSRHHAPALGAGARAIIVAGYARIFRNSVATGEVYPL-LELTDVG-A 188
DB 65 VLVAGPDFGTGSSRHHAPVALMDYGFRAVPSRFRADIFRSGRAGMGTGMEQSDLELL 124
QY 189 WK--ECKTGDVYVDLIANSVFIMHTSGEYKIKPIGDAGPYIEAGIIFAVAR 238
DB 125 WKLMQTPGLELTVNLE-----KQIVTAGDVVISFEVDPYTR 161

RESULT 12

US-09-738-626-5199
; Sequence 5199, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5199
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5199

Query Match 10.3%; Score 130; DB 9; Length 943;
Best Local Similarity 23.1%; Pred. No. 0.00046;

	Matches	54;	Conservative	28;	Mismatches	88;	Indels	64;	Gaps	9;
QY	74	VGDNDIDTQQIIIPAEHLTLVPSKP-----DEFRKIGS-----TAFAQ	109	:	:	:	:	:	:	:
Db	713	LGDSDVTDDHISPPAS--SIKPGTPAAQYLDEGHERHDYNSRGHNHEVMKCTEPANIR	770	:	:	:	:	:	:	:
QY	110	LP-----SAAYPTFPVARG-----EESRY-----ATIVGANPCGSSREHPA	150	:	:	:	:	:	:	:
Db	771	KONOLVDIAGCYTRDTQEGCAPQAFLTYLDASVNYKAAGILPYLVLGKEGTGSSDMAKG	830	:	:	:	:	:	:	:
QY	151	LGAAGARAIYAEGVARIFERNVSATGEVYPLELTDVGAMKECKTGDDVTVYDNLANSFINH	210	:	:	:	:	:	:	:
Db	831	TNLTGRVAIVTESPERLRHSNLIGMG-VVPLQFAPAGESHESIJDGTETFDITGLTALNE	889	:	:	:	:	:	:	:
QY	211	-TSSEKKKLAPIADGAVIE-----AGGFAYARKGMTJMA	246	:	:	:	:	:	:	:
Db	890	GETPRKYAVATKENGVDVEEDAVNRILDPTEGADYYRRHGILLQYVLROMAASK	943	:	:	:	:	:	:	:

[illegible]

```

Db      103 PSFADIEFGNSF-NNQLLPVTLSD 125

      RESULT 14
      US-09-925-302-700
      ; Sequence 700, Application US/09925302
      ; Patent No. US20020044941A1
      ; GENERAL INFORMATION:
      ; APPLICANT: Rosen et al.,
      ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
      ; FILE REFERENCE: PA104
      ; CURRENT APPLICATION NUMBER: US/09/925,302
      ; CURRENT FILING DATE: 2001-08-10
      ; PRIOR APPLICATION NUMBER: PCT/US00/05918
      ; PRIOR FILING DATE: 2000-03-08
      ; PRIOR APPLICATION NUMBER: 60/124,270
      ; PRIOR FILING DATE: 1999-03-12
      ; NUMBER OF SEQ ID NOS: 896
      ; SOFTWARE: PatentIn Ver. 2.0
      ; SEQ ID NO 700
      ; LENGTH: 417
      ; TYPE: PRT
      ; ORGANISM: Homo sapiens
      ;
      US-09-925-302-700

```

```
Query Match          8.8%; Score 112; DB 10; Length 417;
Best Local Similarity 29.7%; Pred. No. 0.0082;
Matches 44; Conservative 14; Mismatches 50; Indels 40; Gaps 6

Qy      74 VGDNIIDPDQIIP-----AEHITLPSKPDXYRKLGS-----FAFAG---- 109 -
       |:::| || | | | | | | | | | | | | | | | | | | | |
Db      199 LGDSVTTHDHSPPGNIARNSPARLYLTNGLPFRFNYSGRGNDAVMARGTFANIRLL 258

Qy      110 ---LPSSAAYPTPEVAFGE-----ESSRY-----AIYGGANFGCGSSREHPALGAA 154
       |:::| | | | | | | | | | | | | | | | | | | | |
Db      259 NREFLNQAQQTIIHLPGSEILDVFDAERYQQAGLPLIVLAGKEYGAGSSRDWMAKPFTLL 318

Qy      155 GARATVAEGGARIFRNNSVATGEVPDE 182
       |:::| | | | | | | | | | | | | | | | | | | | |
Db      319 GKRAVLAEESTERIHRSNLVGMG-VIPLE 345

RESULT 15
US-09-779-307-2
Sequence 2, Application US/09779307
Patent No. US20020137675A1
GENERAL INFORMATION:
APPLICANT: Taupier Jr., Raymond
APPLICANT: Majumder, Kumar
APPLICANT: Vermet, Corine
APPLICANT: Prayaga, Sudhir Das
TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby
FILE REFERENCE: 15966-662 US
CURRENT APPLICATION NUMBER: US/09/779,307
CURRENT FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/180,880
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: 60/181,044
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: 60/181,656
PRIOR FILING DATE: 2000-02-10
PRIOR APPLICATION NUMBER: 60/182,795
PRIOR FILING DATE: 2000-02-15
NUMBER OF SEQ ID NOS: 43
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
US-09-779-307-2

Query Match          8.3%; Score 105.5; DB 10; Length 538;
Best Local Similarity 26.5%; Pred. No. 0.048;
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:06:23 ; Search time 44 Seconds
(without alignments)
544.033 Million cell updates/sec

Title: US-10-027-450-47

Sequence: 1 MAAALSGTAVSTALAPIR.....AGGIFAYARKGTMIASKAA 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710.5	56.1	251	2 H84861	3-isopropylmalate
2	651	51.4	253	2 T47781	3-isopropylmalate
3	627	49.5	256	2 A84862	3-isopropylmalate
4	311.5	24.6	166	2 A72363	3-isopropylmalate
5	272	21.5	161	2 E69328	3-isopropylmalate
6	271	21.4	219	2 E70421	3-isopropylmalate
7	267.5	21.1	168	2 D64459	3-isopropylmalate
8	266	21.0	170	2 E69210	3-isopropylmalate
9	260	20.5	163	2 B97290	3-isopropylmalate
10	258.5	20.4	164	2 E69051	3-isopropylmalate
11	252.5	19.9	162	2 E69051	3-isopropylmalate
12	250.5	19.8	170	2 D64458	3-isopropylmalate
13	249	19.7	165	2 H69459	3-isopropylmalate
14	229.5	18.1	208	2 F75353	3-isopropylmalate
15	227.5	18.0	166	2 D72394	hypothetical prote
16	223.5	17.7	165	2 H90418	hypothetical prote
17	206.5	16.3	163	2 A75160	3-isopropylmalate
18	202.5	16.0	163	2 G71180	probable 3-isoprop
19	192	15.2	659	2 F70453	aconitase - Aquife
20	191.5	15.1	191	2 E36889	probable 3-isoprop
21	189	14.9	202	2 AE1983	3-isopropylmalate
22	188	14.8	215	2 G82364	3-isopropylmalate
23	184	14.5	212	2 F83255	3-isopropylmalate
24	184	14.5	689	2 S26864	3-isopropylmalate
25	183.5	14.5	758	2 T39210	3-isopropylmalate
26	178	14.1	199	2 C69650	3-isopropylmalate
27	174.5	13.8	193	2 A71323	3-isopropylmalate
28	174.5	13.8	194	2 G84031	3-isopropylmalate
29	172	13.6	200	2 S75839	3-isopropylmalate

30	171.5	13.5	202	2 B87273	3-isopropylmalate
31	171.5	13.5	208	2 I39571	probable 3-isoprop
32	170	13.4	213	2 E81128	3-isopropylmalate
33	169.5	13.4	193	2 AG1694	3-isopropylmalate
34	169	13.3	213	2 C81836	probable 3-isoprop
35	168	13.3	200	2 A82071	3-isopropylmalate
36	167.5	13.2	201	2 AE2919	3-isopropylmalate
37	167.5	13.2	216	2 F97693	3-isopropylmalate
38	161.5	12.8	177	2 H75373	3-isopropylmalate
39	159	12.6	190	2 B89998	3-isopropylmalate
40	158.5	12.5	211	2 AB3561	3-isopropylmalate
41	155.5	12.3	201	2 AC0516	3-isopropylmalate
42	155	12.2	201	2 S07306	3-isopropylmalate
43	155	12.2	779	2 S64011	3-isopropylmalate
44	154.5	12.2	201	2 C90638	3-isopropylmalate
45	154.5	12.2	201	2 C85489	isopropylmalate is

ALIGNMENTS

RESULT 1

H84861
3-isopropylmalate dehydratase, small subunit [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84861
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon-Baudry, D.; Niernann, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84861
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-251 <STO>
A:Cross-references: GB:AB02093; NID:93763918; PIDN:AAC64298.1; GSPDB:GN00139*
C:Genetics:
A:Gene: At2g43090
A:Map position: 2

Query Match Best Local Similarity 56.1%; Score 710.5; DB 2; Length 251;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

QY 1 MAAALSGTAVSTALAPIRAPTSFIRRSQTLCHRLSLKCRAGSIVPAAAAAGSS 60
DB 1 MAAALSGTAVSTALAPIRAPTSFIRRSQTLCHRLSLKCRAGSIVPAAAAAGSS 60
QY 61 SPSSA-----VHGECFVGDNDITDQIIPAEHLTVPSKPDYRKLGSFAFAGLPSA 113
DB 57 VTRSAEPQERKTFGLCYVGDNDITDQIIPAEHLTVPSNPEYRKLGSYALVGLP-A 115
QY 114 AVPTTPVAPGESSRAITVGGANFGCCSSRHPAVLGAAGARIVAEGARFFPNV 173
DB 116 SKREKRVQGEKTKYTIISIGENFGCCSSRHPAVLGAAGARIVAEGARFFPNV 175
QY 174 AGEVYPLELTVGAKRECKTGVTVDL--ANSYFINHTSGEKYKLPIDAGPVIAG 231
DB 176 AGEVYPLELTVGAKRECKTGVTVDL--ANSYFINHTSGEKYKLPIDAGPVIAG 234
QY 232 GIFAFAARKGTMIASKAA 248
DB 235 GIFAFAARKGTMIASKAA 251

RESULT 2

T47781
3-isopropylmalate dehydratase-like protein (small subunit) - Arabidopsis thaliana
N:Alternate names: protein F17J16.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47781
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24476
A:Accession: T47781
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <DAN>
A:Cross-references: EMBL:AL163527
A:Experimental source: cultivar Columbia; BAC clone F17J16
C:Genetics:
A:Map position: 3
A:Note: F17J16.40

query Match	51.4%;	Score 651;	DB 2;	Length 253;
Best Local Similarity	59.7%;	Pred. No. 3,1e-46;		
Matches 135;	Conservative 29;	Mismatches 56;	Indels 6;	Gaps 4

Oy 24 SAFLR-RSOLTCRHLSLKCRAGSIYPAAAAAAAAASSSSSSSAVHGEFCFVAGNDITDQ 82
 | :: : | | | :: : | | | | | | | | | |
 Db 29 SPFLQKLSASTIFNKPILTSSATITRVAASSSDSGESITRETFHGLCFVLKNDITDQ 88

QY 83 IIPAEHLTVPSKPDDEYRKLGSFAAGLPSAAYTPFVADGEESRYAIIVGGANFEGGS 1422
 ||||:||||:||||| ||| | ||||| :||:||||| |||||
 Db 89 IIPAEGLTIPSIPEDREKLGSPALNGIPK - FYNERFVVPGEEMSKSYSVIIIGDGNFEGGS 1477

```

Oy      143 SREHPVALGAGARAIVAEGVARIFFRNSVATGEVYPLELTDVGAMKECKTGDDVYVDL 202
         ||||| |||||::||| ||||| |||||::||| ::| |||||::|:
Db      148 SREHPVCLGAGAKAVVAESYARIFFRNCVATGEIIFPLE-SEVRICDECKTGDDVYIEH 206

```

```

0y 203 ---ANSVFHNHSTSEKFKLKPIDGAPVIEACGIFAYARKTGMIAS 245
      :: ||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 207 KEDGSSLLHNHTTRKEYKLKPLDGAPVIDAGGIFAYARKACGIMPS 252

```

RESULT 3

3-isopropylmalate dehydratase, small subunit [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84862

Rilin, X.; Kaul, S.; Rounsley, S.D.; Shea T.P.; Bentol, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, B.; Nierman, W.C.; White, C.O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999

A.TITLE: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A.Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: AB4862
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <STO>
A:Cross-references: GB:AE002093, NID:93763919, PIDD:GN001396
C:Genetics:
A:Gene: At2g43100
A:Map position: 2

Query Match	49.58;	Score 627;	DB 2;	Length 256;
Best Local Similarity	54.28;	Pred. No. 3e-44;		
Matches 136;	Conservative	38;	Mismatches 65;	Indels 12;
			Gaps	6

```
Oy      4 ALSTAVSTAALLAPIRAPTSAFIRSQTLCHRLHSLKCRAG-----SYTPAAAAAAGA    59  
         || :: |::| :|| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     11 ALPCSSTKTSSLATFRSP---FLPRNGSTSLPSSIITSRGTGSPITIIIPRAAESDS   67
```

Oy 60 SSP-SSAVFHGEFVGVGNDITDQILPAEHLLTVPSKPDYEYKLGSFAAGLPSAAPTP 11
:: :: ||| :: | ||||| || : : : ||| :
Db 68 NEALANTTFHGLCYVLKDNIDTDQLIPAGAATFPNQQRDEIAAHALSGLPD-FHKTR 12

[illegible]

QY 179 YPLELTDVGAWKECKTGVDVTVDLANS--VFINTSGKEYKLPIDAGPVIAGGIFAY 236

```

Db      187  PPIE-SEVAVCEECTGTGTIVIEISDSGGLTNHTTGKNNKLSIGDAGPYIDAGGIFAY 245
QY      237  ARRTGMIASKA 247
        || ||| |
Db      246  ARRMGMIPSLA 256

```

RESULT 4

3-isopropylmalate dehydratase, small subunit - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72363
C:Author: D. Clifton, P. A. Gill, S. P. Gwin, M. L. Dodson, R. J. Haft,

NEILSON, A.E.; CLAYTON, R.A.; GILL, S.K.; GRANT, W.D.; GARRETT, M.M.; STEWART, A.M.; COTTON, M.D.; PRATT, W.S.; PHILLIPS, C.A.; RICHARDSON, C.M.
 VOLUMES 300 323-326 1000

Nature 359, 343-349, 1992
 A1:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A1:Reference number: A72200; PMID:99287316; PMID:10360571

A/Accession: A/2303
A/Status: preliminary
A/Molecule type: DNA

A:Cross-references: GB:AE001730, GB:AE000512, NID:g4981062, PIDN:AAD35640.1, PID:g4981062
A:Experimental source: strain MSB8

A;Gene: TM0555

Query match	24.0%	Score 21.5, 25.2	Length 267
Best Local Similarity	41.1%,	Pred. No. 1.6e-18;	
Matches	72;	Conservative	54;
		Mismatches	23;
		Indels	Gaps

```
09 GECIVGNDNIDJDIIFEAHMLDVEGNEBDNNADCELRNCANF-----II
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7 GKAEVVEGDNVNTDEIIPARLYN--TSDPOLAK-----YCMEDARPFGGRD 51
```

```

0Y      SSKRIALLVGGANGCUCSSRENEF VALGSRANALZVALGSRANALZKNO.....- : : :
126      : : : | | | | | | | | | | : : : : : | | | | | : : : : :
Db 52 DIKSIIIVAGENFGCGSSREHAPVAIKAGISCIVIAKTSARIEFRNAINIG---LPIVE 107

```

[illegible]

RESULT 5

3-isopropylmalate dehydratase, small subunit (leuD-1) homolog - *Archaeoglobus fulgidus*

C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C/Accession: E69328

Kirkness, H.F.; Clayton, R.A.; Limu, J.F.; Miller, J.E.; Nelson, R.L.; Rikken, K.; Kienk, H.F.; Clayton, R.A.; Limu, J.F.; Miller, J.E.; Nelson, R.L.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, H.F.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Glodex, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 350, 504-510, 1993.
A: Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Arltach, P.; Kalne, B.P.; Sykes Smith, H.O.; Woese, C.R.; Venter, J.C.

A/Accession: E69328
A/Reference number: A69250; MUID:98049343; PMID:9389475
A/Accession: E69328

n/molecule of polymer
A; Residues: 1-461 <KLE>
A; Cross-references: GB:AE001061; GB:AE000782; NID:g2689384; PIDN:AAB90610.1; PID:g2689384

Query Match	21.5%	Score 272;	DB 2;	Length 161;
Best Local Similarity	37.2%;	Pred. No. 2.7e-15;		
	70.0%	Warwick	59.0%	Totals 20.6%

```

QY      69  GECFVAGNDITDOIIPAEHLTLVPSKPRDEYRKLGSAFAGL-PSAAVPTPFVARGEES 12
        | : ||||| | : : | | | | | | | | | |
Db      2  GRAWKFGDDIDTVILQSKYLI--NEPEE--LAKHVFENLRPEFA-----KEYK 47

```

QY	128	RYAIVGANGCCSSRHHAPVALCAARAIIVGARYIFRRNSVATGEVYPLDTVG	187
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32	
Dd	48	KGDFVAVAGENRGCCSRHHAPLALKATIGIAEVAKSVARIFRRNAINIG----	LRYLECK 103
QY	188	AWKECKQDYYTVDLANSYFINHTSGREYKRLPGD-AGPYIEAGGITAYAR	238
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32	
Dd	104	ETDCKIEDDDEIVNDIEKGYIYNKTKGEYIPNLPDLKTELKELGGLVAFNA	155

RESULT 6

3-Isopropylmalate dehydratase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: E70421
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MIMD:98196666; PMID:9537320
A:Accession: E70421
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-219 <AOFP>
A:Cross-references: GB:AE000738; NID:g2983801; PID:AE000735.1; PID:g2983802; GB:AE000655
A:Experimental source: strain VF5
C:Genetics:
A:Gene: leuJ

Query Match	21.48;	Score 271;	DB 2;	Length 219;
Best Local Similarity	37.58;	Pred. No. 4.7e-15;		
Matches	72;	Conservative	27;	Mismatches 53; Indels 40; Gaps 8;

```

0y 67 FHBCFVNGSNDITDQIIPAEHTLVPSKPDKEYKLGFSFAGLPSAAYPTPVAPGEES 122
      | | : |||:||||||| :| | |
Db 56 FGRGVKFGDNVDTDQIIPARYLN-TSDPYELAK-----HVMEDSEH 96

```

```
0y 127 SRVA-----IIVGANGFCSSSRHAPVALGAGARAIVAEGYRIFFRSVATGEY 179
      :|          |||||:::|          :::|||||:::|
Db 97 PEPAAKEKEDDIIVAGKNFGSSSRHAPIAIKISGVPIVIAKSFARIFFRNAINIG--- 15
```

180 PLELDV-GAMKECKRGDVAVIDLANSFVINTSGKEYL---KPIGDAGPIEAGGIF 23
 154 -LPIVEAPEAVDEIEHGDEIVDLKGVYIKNLRGIGKEYQATKFPKELQD---ILKAGGIM 20

Qy 255 AIAKNIGMIASK 248
 111:: :111
 Db 210 AYAKE--KLASK 219

RESULT 7
D64459
3-Isopropylmalate dehydratase (EC 4.2.1.33) - *Methanococcus jannaschii*

C:\Access

A: Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstein, K.G.; Merrick, J.M.; Glöckner, A.; Larson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. *Science* 273, 1058-1073, 1996
 A: Authors: Kalne, B.P.; Borodovsky, M.; Kleen, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R. *Proc. Natl. Acad. Sci. USA* 93, 1332-1336, 1996
 A: Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
 A: Reference number: A64300; MUID:96337999; PMID:8688087
 A: Accession: D64459
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-168 <BUI>
 A: Cross-references: GB:U67568; GB:L77117; NID:q2826383; PIDN:AA89283.1; PTD:q1591913; T:G1591913
 A: Genetics:
 A: Map position: REV1223059-122253
 A: Keywords: carbon-oxygen lyase; hydro-lyase

Query Match	21.1%;	Score 267.5;	DB 2;	Length 168;
Best Local Similarity	38.3%;	Pred. No. 6.7e-15;		
Matches 69; conservative	29;	Mismatches 57;	Indels 25;	Gaps 9;

```
QY      65 AVFHECEVVGDNIDDDIIPAEHLTVPSKPDRLKLGSFAAGLSAAYPPIFEVAAGE   120  
       :: : |::|::|::| : |::| : | : ||:  
Db     3 SIIRGRVRWKEGNNVDALIPARY--LYYTKPEE--LAQFWMTG-ADPDFPKK-VKPGD   55
```

```

QY      125  ESSRAIIYGGANFGCGSSRREHAPVALCAGARAIVAEGVARIFRNSVATGEVYPLELT 180
          ||||| ||||| ||||| : || :|| :||| :||| : || :
Db      56  -----IIVGKNGFCGSSRREHAPLGKAGAGISCVIAESFARIIFYRNAINVC--LPL-IE 100

```

QY 185 DVGAMKECKTGVDVINDLVANSVFINTHSTSEKYEKLPIGDGP-----VIAGSIFAAYRK 230

DB 107 CKGISAEKVNEDELEVNLLETGEIKNLTITGEVLK----GQKIPEFMMLILEAGSLMPLYKK 160

RESULT 8
E69210
3-isotropon]malate dehydratase. Ieud subunit - Methanobacterium thermoautot

C:\Access

Qiu, D.; Spadatoro, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani
 ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A>Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: *fr*
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: E69210
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-170 <MTH>
 A:Cross-references: GB:AE000860; GB:AE000666; NID:q2621945; PIDB:ABB85327.1; PID:q26
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH829

Query Match	21.0%;	Score 266;	DB 2;	Length 170;	.
Best Local Similarity	39.5%;	Pred. No. 9e-15;			
Matches 70;	Conservative 24;	Mismatches 61;	Indels 22;	Gaps 7	

```
QY      66 VFHGCFFVGVNDITQIIPAEHLITVPSKPDYKKGSGFAFAGL-PSAAYPTFPVAGE 122
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      4  IIRGVWVRFGVNDITMIIPGRYL-----RTFSLDELASHWEGARPEFASQ---VRKGD 55
```

```

09      125 ESSRAIVAGANFGCCSSREHHPVALGACARAIVAEGVARIFRRSVATGEVYPLEYT 18
          ||| ||||| |||| | ||| : ||| : ||| : | : :
Db    56 -----ITVAGRNFCCSSREQPVALKRGAVAIIAESFARIIFYRNAINIG--LPVIMA 10

```

QY 185 DVGAMKECKGIDVWIDPLANSVFINHISREKYEKLPIGD-ASGVIEKGIFAAKRT 240
| | | | | : | | | | | : | | : | |
Db 108 KQDA----DDGDEVSIDLRSGQIRNLITAGSEYRKMPFNDIMLSILEGGGLVNHLYKT 160

RESULT 9
P97290

C: isopropylmalate dehydrogenase, small chain [unpublished] Clostridium acetobutylicum
C: Species: Clostridium acetobutylicum
C: Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C: Accession: B97290
R: Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;
J. Dairy, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A: Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A: Reference number: A96900; M01D:2135925; PMID:21359325
A: Accession: B97290
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-163 <RUB>
A: Cross-references: GB:AE001437; PIDN:PAK81109.1; PID:q15026240; GSPDB:GM00168
A: Experimental source: Clostridium acetobutylicum ATCC824
A: Genetics:

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:05:28 ; Search time 25 Seconds
(without alignments)
413.104 Million cell updates/sec

Title: US-10-027-450-47

Perfect score: 1266
Sequence: 1 MAAALSTAVSTALALPIR.....AGIPAVARKTGMASKAA 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt-40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	311.5	24.6	166	LEUD_THEMA	Q9W225 thermotoga
2	272	21.5	161	LE23_ARCEU	O29626 archaeoglob
3	271	21.4	168	LEUD_AOUAE	O67399 aquifex ae
4	267.5	21.1	168	LE24_METUA	O58673 methanococ
5	266	21.0	170	LE23_METTH	O26917 methanobact
6	252.5	19.9	162	LE24_METTH	O27440 methanobact
7	250.5	19.8	170	LE23_METUA	O38667 methanococ
8	249	19.7	165	LE24_ARCEU	O28513 archaeoglob
9	202.5	16.0	163	LEUD_PYRHO	O53393 pyrococcus
10	191.5	15.1	191	LEUD_LACLA	O02144 lactococcus
11	189	14.9	215	LEUD_AZOVI	P66196 azotobacter
12	184	14.5	744	LEUD_PHYBL	P18250 phycomyces
13	183.5	14.5	758	LEUD_SCHFO	O14289 schizosach
14	180	14.2	750	LEUD_RHINI	P53811 rhizopus ni
15	178	14.1	199	LEUD_BACSU	P94568 bacillus ni
16	176	13.9	773	LEUD_USTMA	P49601 ustilago ma
17	172	13.6	200	LEUD_SYNY3	P74207 synechocyst
18	171.5	13.5	208	LEUD_ALCEU	O44022 alcaligenes
19	171	13.5	201	LEUD_THERH	O92nd2 thermus the
20	165	13.0	770	LEUD_CANNA	O00464 candida mal
21	164.5	13.0	755	LEUD_RHIPU	P55251 rhizomucor
22	155	12.2	779	LEUD_SALTY	P04787 salmonella
23	155	12.2	779	LEUD_YEAST	P07264 saccharomyc
24	153.5	12.1	200	LEUD_ECOLI	P30126 escherichia
25	151	11.9	201	LEUD_MAEIN	P44368 haemophilus
26	144.5	11.4	693	LEUD_YEAST	P43637 saccharomyc
27	139.5	11.0	197	LEUD_STRICO	O86535 streptomyce
28	138.5	10.9	137	LEUD_NEIRA	P50181 neisseria i
29	133.5	10.5	100	LEUD_CLOPA	P31960 clostridium
30	133	10.4	207	LEUD_BUCPP	O92ez5 buchnera ap
31	131.5	10.4	887	ACOC_CABEL	O25500 caenorhabdi
32	129	10.2	775	LEUD_MMENT	O94412 emericella
33	128	10.1	198	LEUD_MYCLE	O33124 mycobacteri

34	128	10.1	198	LEUD_MYCTU	O53236 mycobacteri
35	127	10.0	207	LEUD_BUCAI	P56935 buchnera ap
36	126	10.0	208	LEUD_BUCDN	O85073 buchnera ap
37	125	9.9	961	ACON_MYCAV	O08451 mycobacteri
38	122	9.6	207	LEUD_BUCAP	O85066 buchnera ap
39	122	9.6	207	LEUD_BUCRP	P48574 buchnera ap
40	120.5	9.5	205	LEUD_BUCTS	O31294 buchnera ap
41	118	9.3	891	ACON_LEGPN	P37032 legione11a
42	116.5	9.2	780	ACON_BOVIN	P20004 bos taurus
43	116	9.2	878	ACON_RICCN	O92950 rickettsia
44	115.5	9.1	780	ACON_HUMAN	O99798 homo sapien
45	114.5	9.0	781	ACON_PIC	P16276 sus scrofa

ALIGNMENTS

RESULT 1	ID	LEUD_THEMA	STANDARD:	PRT:	166 AA.
AC	Q9W225	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	3-isopropylmalate dehydratase small subunit (PC 4.2.1.33)				
DE	(Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).				
GN	LEUD OR TW0555.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogae (class); Thermotogales;				
OC	Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109;				
RX	MEDLINE=99287316; PubMed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RL	"Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";				
CC	Nature 399:323-329(1999).				
CC	-1- CARBATIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE + H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-4-METHYL-3-CARBOXY-PENTANONE).				
CC	-1- PATHWAY: leucine biosynthesis; second step.				
CC	-1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	or send an email to license@isb-sib.ch .				
DR	EMBL: AE001750; AAD35640.1; -.				
DR	TIGR: TM0555; -.				
DR	InterPro: IPR000573; Aconitase_C.				
DR	Pfam: PF00694; Aconitase_C; 1.				
KW	leucine biosynthesis; lyase; Complete proteome.				
SO	SEQUENCE 166 AA; 18451 MW; B556DCBB6B0D22F.CRC64;				
QY	69 GECFVAGNIDTDIIPAEHLTVPSKPDYRKGSFAAGLPSAAATPTPVAGG---EE 125				
DB	7 GKVFVGGDNVTDEIPARYLN--TSDFOELAK-----YCMEDARGGGRD 51				
	Query Match	24.6%;	Score 311.5;	DB 1;	Length 166;
	Best Local Similarity	41.1%;	Pred. No. 1e-18;		
	Matches	72;	Conservative	26;	Mismatches 54;
				Indels	23;
				Gaps	5;

[illegible]

RESULT 2	LE23_ARCFU	STANDARD:	PRT:	161 AA.
ID	LE23_ARCFU			
AC	029626;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)			
DE	(isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).			
GN	AF06529.			
OS	Archaeoglobus fulgidus.			
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;			
OC	Archaeoglobaceae; Archaeoglobus.			
OX	NCBI_TaxID=2234;			

RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleih H.-P., Clayton R.A., Tombl J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gilm J., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodak A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Fujie B.P., Sykes S.M.,
RA Sadow P.M., D'Andrea K.P., Bowman C., Kallal C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H₂O (ALSO CATALYSES 2-ISOPROPYLMALATE + H₂O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: leucine biosynthesis, second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).

CC This SWI-S-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE001061; LAMB90610.1; -
DR HSSP: P20004; LAMT.
DR TIGR: AF0629; -
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00694; Aconitase_C.1.
KW leucine biosynthesis; lyase; Complete proteome.
KW SEQUENCE 161 AA; 18137 MW; 6DBE80A2AE0B16A7 CRC64;

Query Match 21.5%; Score 272; DB 1; Length 161;
Best Local Similarity 37.2%; Pred. No. 1.6e-15;
Matches 64; Conservative 30; Mismatches 58; Indels 20; Gaps 6;

```

QY      69  GCGCFVGNQIDTDQIIPAEHNTLTPKSPDDYKRLKLSFAAGL--PSAAALPTFPVARGESS 127
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      2  GRAMFEGDDITDVLIOCKYLV--NPEPE---LAKHVENLRLPEFA-----KEVK 47
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY     128  RAILIVGAGNFGCGSSRKHAFVVLGAAGAAIYAEGYAIRFRNSVATGEVYPLELTDVG 187
      1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

D0 48 KGFVAVGENFGSSRREHAPLAKA'GIEAVIAKSYARIFFRNAINIG----LRYLECK 103

OY 108 AMKECTEDVVTVDIANSVFINTHTSGEYKKIPLGD-AGPYIEANGJFEAFAYR 238
 :: :: :: | - || :|| :: | :: ||| :: :| :

D6 104 ETDKTEDDELEVDEKYEVLYNNKGGEYPINLPDLKLLELGGGLVERPK 155

RESULT 3			
LEUD_AQUAE			
ID	LEUD_AQUAE	STANDARD:	PRT; 168 A.
AC	067399;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	1-isopropylmalate dehydratase small subunit (EC 4.2.1.33)		
DE	(Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).		
GN	LEUD OR Aq_1398.		
OS	Aquifex aeolicus.		
OC	Bacteriia: Aquificae; Aquificae (class); Aquificales; Aquificaceae;		
CC	Aquifex.		
OX	NCBI_TaxID=63363;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RX MEDLINE=981196666; PubMed=95373220;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Saason R.V.:
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus";
 RL Nature 392:353-358(1998).
 RL
 RL Nature 392:353-358(1998).
 CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
 CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
 CC 4-METHYL-3-CARBOXPENTANONE).
 CC
 CC -1- PATHWAY: Leucine biosynthesis; second step.
 CC
 CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
 CC
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AE000738; AAC07359.1; ALT INIT.
 DR InterPro: IPR000573; Aconitase_C.
 DR Pfam: PF00664; Aconitase_C.1.
 CC
 CC Leucine biosynthesis; Lyase; Complete proteome.
 KW
 SO SOURCE 168 AA; 18786 MW; 25A34FE7FCA09F3 CRC64;

Query Match	21.48;	Score 271;	DB 1;	Length 168;
Best Local Similarity	37.58;	Pred. No. 2.1e-15;		
Matches	72;	Conservative	27;	Mismatches 53; Indels 40; Gaps 8.

```
QY      67 FHGCEVVGNIIDTQIIIPAEHLTLVSPKDEYKLGSEFAAGLPSAAYPTTFVAAGEES 120
        | : |||:|||||| : | | | | | | | | | | | | | | | | | | | | |
Db      5 FRGRVWKFGENVDTQIIPARLN--TSDPYLAK-----HVMEDSEH 45
```

QY 127 SRA-----IIIGCANFGCGSSREHAPALACAGARATVAEGVARIIFRNSVATGEVY 179
| | | | |
Db 46 PEFIAKHCKGCDITLVAKNFGSGSSREHAIPAIKRYGVPAIVAKSEFRPIFFRNALIG-- 107
| | | | |
QY 180 PLFLTDV-GAMKECTGDVVTVLANSVFINTHSGREYL---KRISDAGVIEAGIIF 234
| | | | |
Db 103 -LEIVAEPAVDVEIHGHGEIVDLERKVINLRTRGEEVATFEPPELD---ILKAGIM 156
| | | | |

QY	235	AYARKTGMIASK	246
	::	::	
Db	159	AYAKE--KLASK	168

RESULT 4

```

LE24_METTA          STANDARD:          PRT:          168 AA.
ID LE24_METTA
AC 058673:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN MJ1277.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=6688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleink H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U67568; AAB9283.1; -.
CC DR TIGR: MJ1277; -.
CC DR InterPro: IPR000573; Aconitase_C.
CC DR Pfam: PF00694; Aconitase_C; 1.
CC KM Leucine biosynthesis; Lyase; Complete proteome.
CC SQ SEQUENCE 168 AA; 18377 MW; A53C2ACAB83B6C07 CRC64;

Query Match 21.1%; Score 267.5; DB 1; Length 168;
Best Local Similarity 38.3%; Pred. No. 4e-15;
Matches 69; Conservative 29; Mismatches 57; Indels 25; Gaps 9;

QY 65 AVFGCEFYVGNIDTDQIIPAEHLTVPSKPDVEYRKLGSAFAAGLPSAAYPTFFVAPGE 124
DB 3 SIIRGVWVRGMDVDDMIIPGRYL-----RFFSDELASIMVEGARREFPSQ---VRKGD 55
QY 125 ESSRAIIVYGANFGCGSSREHAPYALGAGARAIYAEGARIRIFRNSVATGEYPLELT 184
DB 56 -----IIVAGNFCCGSSREHAPYALGAGARAIYAEGARIRIFRNSVATGEYPLELT 106
QY 185 DVGAMKECKTGDPVYVDLANSVFNIHTSGKEYKLPIDGAP-----VIEAGIGFAVARK 239
DB 107 CAGISKRVNKGDELEVNLETGTEIKNLITGVEYLK-----GQKPEPMETLEAGGIMPLIKK 162

RESULT 5
LE24_METTH          STANDARD:          PRT:          170 AA.
ID LE24_METTH
AC 026917;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

```

DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN MTH829.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Delta H.
RX MEDLINE=96037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiweni N., Carnuso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Piotrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
CC J. Bacteriol. 179:7135-7155(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE000860; AAB85327.1; -.
CC DR InterPro: IPR000573; Aconitase_C.
CC DR Pfam: PF00694; Aconitase_C; 1.
CC KM Leucine biosynthesis; Lyase; Complete proteome.
CC SQ SEQUENCE 170 AA; 18651 MW; BC07146AF7585215 CRC64;

Query Match 21.0%; Score 266; DB 1; Length 170;
Best Local Similarity 39.5%; Pred. No. 5.4e-15;
Matches 70; Conservative 24; Mismatches 61; Indels 22; Gaps 7;

QY 66 VFHGCIFYVGNIDTDQIIPAEHLTVPSKPDVEYRKLGSAFAAGLPSAAYPTFFVAPGE 124
DB 4 IIRGVWVRGMDVDDMIIPGRYL-----RFFSDELASIMVEGARREFPSQ---VRKGD 55
QY 125 ESSRAIIVYGANFGCGSSREHAPYALGAGARAIYAEGARIRIFRNSVATGEYPLELT 184
DB 56 -----IIVAGNFCCGSSREHAPYALGAGARAIYAEGARIRIFRNSVATGEYPLELT 107
QY 185 DVGAMKECKTGDPVYVDLANSVFNIHTSGKEYKLPIDGAP-----AGPYIEAGIGFAVARK 240
DB 108 KYDA-----DDGDEIVSIDLRSGQIRNLITGSGEYRKPFNDYLSLDEGGGLVNHKTK 160

RESULT 6
LE24_METTH          STANDARD:          PRT:          162 AA.
ID LE24_METTH
AC 027440;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN MTH1387.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-Delta H.
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Hashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoan L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadifora R., Vicare R., Wang Y., Wleziowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pletrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7153(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC -----
CC EMBL: AE000901; AAB85864.1; -.
CC Interpro: IPR000573; Aconitase_C.
CC Pfam: PF00694; Aconitase_C; 1.
CC Leucine biosynthesis; Lyase; Complete proteome.
CC KW SEQUENCE 162 AA; 17956 MW; FF6978208DE0F5B1 CRC64;
SQ
Query Match 19.9%; Score 252.5; DB 1; Length 162;
Best Local Similarity 37.9%; Pred. No. 6.4e-14;
Matches 66; Conservative 26; Mismatches 63; Indels 19; Gaps 8;
QY 69 GECVGVGNDITDQIIIAEHITLVPSPDEYRKLGSAFAGLPSSAATPTFFVAPGESSR 128
DB 3 GKVKFPDDVDVTDIIIGRVLVW----RDPEKLRHEVMGL-DPEPSK-VKPGD---- 51
QY 129 YAIIVGANFGCGSSREHAPVALGAAGARAIVAGYARIFFRNSVATGEVYPLELTDVGA 188
DB 52 --FIVAGKNECGSSREHAPVALKAGIAIVANESFARIFYRNAINNG--IPL-LEAPGI 106
QY 189 WKECKTGDVYVTDIANSVFINTSGKEYKLPICD-AGPYIENAGCIPAYARKTG 241
DB 107 TEKINEGDEIVDDIRGVYLR--GDDEFPFKLPDPFVVEILKGLPIYLKKGK 158

```

```

RESULT 7
LE23_METUA STANDARD; PRT; 170 AA.
ID LE23_METUA
AC Q58667;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN MJ1271.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bolt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kellavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

```

```

RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.M., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleen H.-P., Fraser K.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67568; AAB99277.1; -.
CC HSP: P16276; TACN.
CC TIGR: MJ1271.
CC Interpro: IPR000573; Aconitase_C.
CC Pfam: PF00694; Aconitase_C; 1.
CC Leucine biosynthesis; Lyase; Complete proteome.
CC KW SEQUENCE 170 AA; 18665 MW; 58146C47877ECB7 CRC64;
SQ
Query Match 19.8%; Score 250.5; DB 1; Length 170;
Best Local Similarity 35.2%; Pred. No. 9.9e-14;
Matches 64; Conservative 30; Mismatches 69; Indels 19; Gaps 7;
QY 66 VHGCEVGVGNDITDQIIIAEHITLVPSPDEYRKLGSAFAGLPSSAATPTFFVAPGEE 125
DB 2 IIRGKRAHFGDDVDVTDIIIPRYLTT----DPL-ELASHCMAGI-DENFPKK-VKESD- 53
QY 126 SSRVAIVGANFGCGSSREHAPVALGAAGARAIVAGYARIFFRNSVATGEVYPLELTD 185
DB 54 ----VIVAGNFGCGSSREHAPVALKAGIAIVANESFARIFYRNAINVGLPIPIANTD 108
QY 186 VGAKKECKTGDVYVTDIANSVFINTSGKEYKLPICDAGPYIENAGCIPAYARKTGMA 244
DB 109 -----EIKNDGIVIEDLKEELIVTNNKKTICKCEPKLERITLAAGLVNLKRRKLIQ 163
QY 245 SK 246
DB 164 SK 165

```

```

RESULT 8
LE24_ARCFU STANDARD; PRT; 165 AA.
ID LE24_ARCFU
AC Q28513;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN AF1761.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleen H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.A., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kellavage A.R., Graham D.E., Kyriakides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,

```



```

CC      4-METHYL-3-CARBOXYPENTANONE).
CC      -1- PATHWAY: Leucine biosynthesis; second step.
CC      -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL; U02974; AAB81916.1; -.
CC      DR EMBL; AE006354; AAK05319.1; -.
CC      DR PIR; S35135; S35135.
CC      DR InterPro: IPR000573; Aconitase_C.
CC      DR InterPro: IPR004431; Leud.
CC      DR Pfam; PF00694; Aconitase_C; 1.
CC      DR TIGRFAMs; TIGR00171; Leud; 1.
CC      KW Leucine biosynthesis; Lyase; Complete proteome.
CC      SEQUENCE 191 AA; 21849 MW; C691BC6AEFE4E964 CRC64;
SQ
Query Match 15.1%; Score 191.5; DB 1; Length 191;
Best Local Similarity 29.0%; Pred. No. 7.3e-09;
Matches 58; Conservative 35; Mismatches 70; Indels 37; Gaps 7;
OY 58 GSSPSSAVFGECEFWGDDIDTDQIIPAEHLITVPSKDEYRKIGSFAGALPSAAYPT 117
Db 9 GTSPV-----VMDNIDTDQIIPKPLKAIIDKK--GFGKNLFYEMRYLKDYDENP 56
OY 118 PFVAPGESSRYAIIIGGANGCGSSRHHAVYALGAGARAIIVAGYARIFRRSVATGE 177
Db 57 DFLINAPRYKKAALLISDNGSSSRHHAAWALSDYGFRAIIAGSYSDIFYNALKNG- 115
OY 178 VYPL-----ELTDVGAMKECKTGADVTVDLANSYFINNTSGKEYLKP----- 221
Db 116 LLPKQPREVNLQILKLSQGE-----ITIDPRQLITSLGDPHFHEDIPDWKDLING 169
OY 222 -GDAGPVIE-AGGIFAYARK 239
Db 170 LDDIGITQLQYEBEIAISYAEOK 189
RESULT 11
LEUD_AZOVI STANDARD: PRT: 215 AA.
ID LEUD_AZOVI
AC P96196;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-Isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
DE LEUD.
GN Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=354;
RX STRAIN=UM;
RX MEDLINE=97261869; PubMed=9108283;
RA Manna A.C., Das H.K.;
RT "Characterization and mutagenesis of the leucine biosynthetic genes
RT of Azotobacter vinelandii: an analysis of the rarity of amino acid
RT auxotrophs.";
RL Mol. Gen. Genet. 254:207-217(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
CC H(2)O (ALSO CATALYSES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
CC 4-METHYL-3-CARBOXPENTANONE).
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

DR EMBL; Y11280; CAA72150.1; -

DR InterPro; IPR000573; Aconitase_C.

DR InterPro; IPR004431; Leud.

DR Pfam; PF00694; Aconitase_C; 1.

DR TIGRFAMs; TIGR00171; leud; 1.

KW Leucine biosynthesis; Lyase.

SO SEQUENCE 215 AA; 2413 MW; 8E7CC71322800588 CRC64;

Query Match 14.9%; Score 189; DB 1; Length 215;

Best Local Similarity 35.2%; Pred. No. 1.3e-08;

Matches 56; Conservative 26; Mismatches 59; Indels 18; Gaps 8;

Db 136 DELPROCESTEGYRLTVDLAQT-VTRPDGKALSFEDIP 173

QY 189 ---WKECKT--GDVYTVDLANSVFINHNSGK--EKRLKP 220

Db 77 ASVLARENFEGCGSSREHAPALDXYGFFRTVAPSPADFNFNNSRKNG-LTPIILLPEAVY 135

QY 129 VAIIVGGANFEGCGSSREHAPVALGAAGARIVAEGARIFFRNSVATGSEVPLELTDYGA 188

Db 77 NIDTDQIIPAEHLTLV-----PSKRPDEYKLGSAFAFGLPSEAAVP--TPFYAPGEESR 128

QY 18 NVDTDQIIPKQFLSIKRTGFGPNLDEWKRYL-DVGQPGQDCSARPLNGLFVNLPRYG 76

Db 129 VAIIVGGANFEGCGSSREHAPVALGAAGARIVAEGARIFFRNSVATGSEVPLELTDYGA 188

QY 77 ASVLARENFEGCGSSREHAPALDXYGFFRTVAPSPADFNFNNSRKNG-LTPIILLPEAVY 135

QY 189 ---WKECKT--GDVYTVDLANSVFINHNSGK--EKRLKP 220

Db 136 DELPROCESTEGYRLTVDLAQT-VTRPDGKALSFEDIP 173

RESULT 12

LEU2_PHYBL STANDARD: PRT; 744 AA.

ID LEU2_PHYBL

AC P18250:

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 3-Isopropylmalate dehydratase (EC 4.2.1.33) (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).

DE GN LEU1.

OS Physcomyces blakesleeanus.

OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae; Physcomyces.

OC NCBI_TaxID=4837;

OX [1]

RN SEQUENCE FROM N.A.

RP STRAIN=NRRL 1555;

RC MEDLINE=90356419; PubMed=2388845;

RX Iturriaga E.A., Diaz-Minguez J.M., Benito E.P., Alvarez M.I., Esalva A.P.;

RA Nucleotide sequence of the physcomyces blakesleeanus leu1 gene.";

RL Nucleic Acids Res. 18:4612-4612(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL 1555;

RX MEDLINE=92224296; PubMed=1563047;

RA Iturriaga E.A., Diaz-Minguez J.M., Benito E.P., Alvarez M.I., Esalva A.P.;

RA "Heterologous transformation of Mucor circinelloides with the physcomyces blakesleeanus leu1 gene.";

RT Curr. Genet. 21:215-223(1992).

RL [3]

RN IDENTIFICATION OF PROBABLE FRAMESHIFT.

RP Gibson T.J.;

RA Unpublished observations (MAR-1996).

CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate and 3-isopropylmalate, via the formation of 2-isopropylmaleate.

CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmaleate + H(2)O.

CC -1- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-isopropylmalate.

```

CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -1- CAUTION: Ref.1 and Ref.2 sequences differ from that shown due to a
CC   frameshift in position 676
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53090; CAA37257.1; ALT_FRAME.
CC PIR: S10998; S10998.
CC PIR: S26864; S26864.
CC InterPro: IPR000573; Aconitase_C.
CC InterPro: IPR001030; Aconitase_N.
CC InterPro: IPR004430; Aconitase_N.
CC InterPro: IPR004431; Leu.
CC Pfam: PF00330; aconitase_1.
CC Pfam: PF00694; Aconitase_C.1.
CC PRINTS: PR00415; ACONITASE.
CC ProDom: PD000511; Aconitase_N.1.
CC TIGRFS: TIGR00170; leuc.1.
CC TIGRFS: TIGR00171; leuc.1.
CC PROSITE: PS00450; ACONITASE_1.1.
CC PROSITE: PS01244; ACONITASE_2.1.
CC Leucine biosynthesis: Lyase: Iron-sulfur; 4Fe-4S.
CC FT METAL 341 341 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 401 401 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 404 404 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SQ SEQUENCE 744 AA; 80833 MW; 9E5A930891D3745C CRC64;

Query Match 14.5%; Score 184; DB 1; Length 744;
Best Local Similarity 38.0%; Pred. No. 1.4e-07;
Matches 57; Conservative 15; Mismatches 62; Indels 16; Gaps 8;

OY 42 CRRAGSYVPAAMAAAGSSPSSAVFHGECFVG-DNIDTQIIPAEHLTVPSKPDEYR 100
DB 499 CRRAGSYVPAAMAAAGSSPSSAVFHGECFVG-DNIDTQIIPAEHLTVPSKPDEYR 100
OY 101 KUGSPAFLG--PSAAVTP-FV--ADGESSRYAIIIGVANGCGSSREHAPVALGAA 154
DB 552 GUGSALFYGLRYPATGAKEPFLVNPQYRRSK-ILVCTGPNMGCGSSREHAPVALGAA 154
OY 155 GARAIYAEYARIFFRNSVATGEYVPLELT 184
DB 611 GIRCIITATSFADIFRNCKKNG-MLPITLS 639

RESULT 13
LEU2.SCHPO STANDARD: PRT; 758 AA.
AC 014289:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate
DE 1 isomerase) (Alpha-IPM isomerase) (IPMI).
GN SPAC9E9.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; Pubmed=11859360;
RA Wood V., Guilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

```

```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jajels K.,
RA James K., Jones L., Jones M., Leather S., Mcdonald S., Mclean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Ocell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton K., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,
RA Willems I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler E., Dreano S., Gault R., Lelaure V., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Gault R., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovskii G.V., Usero D., Barrell B.G., Nurse P.,
RA The genome sequence of Schizosaccharomyces pombe.
RA Nature 415:871-880(2002).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC and 3-isopropylmalate, via the formation of 2-isopropylmalate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
CC H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-
CC isopropylmalate.
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z99262; CAB16402.1; -.
CC InterPro: IPR000573; Aconitase_C.
CC InterPro: IPR001030; Aconitase_N.
CC InterPro: IPR004430; Leuc.
CC InterPro: IPR004431; Leuc.
CC Pfam: PF00330; aconitase_1.
CC Pfam: PF00694; Aconitase_C.1.
CC PRINTS: PR00415; ACONITASE.
CC ProDom: PD000511; Aconitase_N.1.
CC TIGRFS: TIGR00170; leuc.1.
CC TIGRFS: TIGR00171; leuc.1.
CC PROSITE: PS00450; ACONITASE_1.1.
CC PROSITE: PS01244; ACONITASE_2.1.
CC Leucine biosynthesis: Lyase: Iron-sulfur; 4Fe-4S.
CC FT METAL 359 359 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 420 420 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 423 423 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SQ SEQUENCE 758 AA; 82782 MW; CE78C36828380E47 CRC64;

Query Match 14.5%; Score 183.5; DB 1; Length 758;
Best Local Similarity 32.1%; Pred. No. 1.6e-07;
Matches 60; Conservative 19; Mismatches 77; Indels 31; Gaps 6;

OY 53 AAAAAAGSSPSSAVFHGECFVG-----DNIDTQIIPAEHLTVPSKPDEYRKGSF 105
DB 523 ATNVAGSVSSGSGIPKFTYVGEIAPLPMANVDTKTIIPKQFL-----TKIKTGTGGGF 577
OY 106 AFAGLPSAAYPTFPVAPGESGRY-----AIIYGANFCGSSREHAPVALGAAG 155
DB 578 AF-----YETRYADAGKEIPDFVLANEPRKATVLAHNRFCGSSREHAPVALNDGF 630
OY 156 ARAIYAEYARIFFRNSVATGEV-YPLELTVGAMKECKTGDV-VTVLANSVYFINHTSG 213

```

DB 631 IRTVIAESFADIFFNNCKRNGMLPIPIPIEVDYNDMAKRAENQVAFSDVLNQTITGDKQ 690
QY 214 KEYCLKP 220
DB 691 VKFVEP 697
RESULT 14
LEU2_RHINI STANDARD; PRT; 750 AA.
AC P55811;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate
isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEU1.
OS Rhizopus niveus.
CC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
CC Rhizopus
CC NCBI_TaxID=4844;
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-Yamazaki / IFO 4810;
RX MEDLINE=97056764; PubMed=8901103;
RA Takaya N., Yanai K., Horuchi H., Ohta A., Takagi M.;
RT "Cloning and characterization of the Rhizopus niveus leu1 gene and
its use for homologous transformation.";
RL Biosci. Biotechnol. Biochem. 60:448-452(1996).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
and 3-isopropylmalate, via the formation of 2-isopropylmalate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-
isopropylmalate.
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: Monomer (by similarity).
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D63833; BA009893.1; -
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR InterPro; IPR004430; Leuc.
DR InterPro; IPR004431; Leud.
DR Pfam; PF00330; aconitase; 1.
DR Pfam; PF00694; aconitase; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR00170; leuc; 1.
DR TIGRFAMs; TIGR00171; leud; 1.
DR PROSITE; PS00450; ACONITASE_1; 1.
DR PROSITE; PS01244; ACONITASE_2; 1.
KM Leucine biosynthesis; Lyase; iron-sulfur; 4Fe-4S.
FT METAL 353 353
FT METAL 413 413
FT METAL 416 416
SQ SEQUENCE 750 AA; 61308 MW; 7D7BE28304E26A83 CRC64;
Query Match 14.28 Score 180; DB 1; Length 750;
Best Local Similarity 34.58; Pred. No. 3e-07;
Matches 59; Conservativity 20; Mismatches 72; Indels 20; Gaps 8;
QY 50 PAAAAAAGSSSPSSAVHHEGEFVG-DNIPDQIIPHEHILTVSKDEVRKIGSFAPA 108
||:|||||:|:|:|||||||:|:|:|||||

DB 515 PASDSSSG-GMPAFITLKGYAAPIDISNIDDMIIKPOFL-----KTIKRTGLSALFY 568
QY 109 GL---PSAAYPPPEYVAGESSRYA--IIVGANFEGCGSSREHAPVALGACARIVAE 163
DB 569 SLRFDPQTGAENPAFLAKERTFROARILVCTGPNRCGSSREHAPWAFNDFGRIILAPS 628
QY 164 YARIEFRNSVANGGEYPL-----ELTDVGAMKECKTGVDVTVDLANSYRN 209
DB 629 FADIFFNNCKRNG-MLPIVLPQAQLFAINA--EAQKGEVEVDLVOQIVRN 676
RESULT 15
LEUD_BACSU STANDARD; PRT; 199 AA.
AC P94568;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)
DE (isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
GN LEUD.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC NCBI_TaxID=1423;
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RX MEDLINE=97124191; PubMed=8969504;
RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emerson P.T., Harwood C.R.;
RT "The dnaB-chromosome (256 degrees-240 degrees) region of the Bacillus
subtilis chromosome containing genes responsible for stress
responses, the utilization of plant cell walls and primary
metabolism.";
RT Microbiology 142:3067-3078(1996).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Deniot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlin K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Pario V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schoefer R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seir S.J., Serrro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takenishi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Wellenzeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: 3-ISOPROPYLMALATE = 2-ISOPROPYLMALATE +
H(2)O (ALSO CATALYZES 2-ISOPROPYLMALATE + H(2)O = 3-HYDROXY-
4-METHYL-3-CARBOXPENTANONE).


```
CC -!- PATHWAY: leucine biosynthesis; second step.
CC -!- SUBUNIT: CONSISTS OF TWO DIFFERENT SUBUNITS: LEUC AND LEUD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 275208; CAA9534.1; -.
DR EMBL: 299118; CAB14785.1; -.
DR Subtilist; BG11950; leud.
DR InterPro; IPR000573; Aconitase_C.
DR Pfam; PF00694; Aconitase_C; 1.
DR TIGRFAMs; TIGR00171; leud; 1.
KW leucine biosynthesis; lyase; Complete proteome.
SQ SEQUENCE 199 AA: 23022 MW: 99ADPE12A0D5D9D3 CRC64;

Query Match 14.1%; Score 178; DB 1; Length 199;
Best Local Similarity 34.2%; Pred. NO. 9.7e-08;
Matches 53; Conservative 22; Mismatches 48; Indels 32; Gaps 8;

QY 77 NIDTDQIIPAEHLTVPSKPDKEYRKLGSFAPAGLPSAAYPTPFVAPGESSRY----- 129
DB 18 NVDTDQIIPKQFL-----KRIERTGYGRFAF-----FDWKRIDANGEPNPEFELNDPVY 65

QY 130 ---AIIVGANFGCGSSREHAPALGAGARAIVAGYARIFFRNSVATGVEVPLELDV 186
DB 66 QGASIIIDAGENFGCGSSREHAPALDYGFKIITIASFADIFHQWCFKNG-MLPIRM-py 123

QY 187 GAMKECKTGDV-----VTVDLANSVFINTSGKE 215
DB 124 DNMKQL-VGQYENOSIQMTVDLENQL-IHDSGNQ 156
```

Search completed: March 31, 2003, 09:08:09
Job time : 26 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 09:05:48 ; Search time 33 Seconds

(without alignments)
1554.719 Million cell updates/sec

Title: US-10-027-450-47

Perfect score: 1266
Sequence: 1 MAAALSTAVSTAALLAPR.....AGIFPAVKRTGMISKAAA 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	990.5	78.2	257	10 Q949D1	Q949d1 oryza sativ
2	710.5	56.1	251	10 Q9ZM85	Q9ZM85 arabidopsis
3	651	51.4	253	10 Q91Y77	Q91Y77 arabidopsis
4	627	49.5	256	10 Q9ZM84	Q9ZM84 arabidopsis
5	265.5	21.0	164	17 Q8U240	Q8U240 pyrococcus
6	260	20.5	163	16 Q97EE1	Q97EE1 clostridium
7	258.5	20.4	170	17 Q9U206	Q9U206 pyrococcus
8	254.5	20.1	164	17 Q8TX94	Q8TX94 methanopyru
9	252.5	19.9	161	16 Q8RDX1	Q8RDX1 thermoaer
10	252	19.9	162	17 Q8TU71	Q8TU71 methanosarc
11	231	18.2	161	17 Q8ZM36	Q8ZM36 pyrobaculum
12	229.5	18.1	208	16 Q9RT10	Q9RT10 delinococcus
13	227.5	18.0	164	17 Q8TRF7	Q8TRF7 methanosarc
14	227.5	18.0	166	16 Q9WY08	Q9WY08 thermotoga
15	225.5	17.8	166	17 Q8TJM9	Q8TJM9 methanosarc
16	223.5	17.7	165	17 Q97YX3	Q97YX3 sulfolobus

17	222	17.5	168	17 Q974Q9	Q974q9 sulfolobus
18	212	16.7	168	17 Q8TW31	Q8TW31 methanopyru
19	210	16.6	164	17 Q8U0B9	Q8U0B9 pyrococcus
20	206.5	16.3	163	17 Q9V1I9	Q9V1I9 pyrococcus
21	202	16.0	163	2 Q92ND9	Q92ND9 thermus the
22	192	15.2	659	16 Q67556	Q67556 aquifex aeo
23	189	14.9	202	16 Q8YX03	Q8YX03 anabaena sp
24	188	14.8	215	16 Q9PAX1	Q9PAX1 xylella fas
25	184	14.5	212	16 Q9HZ44	Q9HZ44 pseudomonas
26	182.5	14.4	216	16 Q8XXX4	Q8XXX4 ralsstonia s
27	179.5	14.2	711	3 Q9P3Y5	Q9P3Y5 yarrowia li
28	178	14.1	196	2 Q9ATM2	Q9ATM2 streptococc
29	176.5	13.9	201	16 Q92LAI	Q92LAI rhizobium m
30	174.5	13.8	193	16 Q8Y5R6	Q8Y5R6 listeria m
31	174.5	13.8	194	16 Q9K8F1	Q9K8F1 listeria m
32	172.5	13.6	641	16 Q8RCF8	Q8RCF8 bacillus ha
33	171.5	13.5	202	16 Q9A8N1	Q9A8N1 thermoaer
34	170	13.4	201	2 Q8RP98	Q8RP98 cauliobacter
35	170	13.4	213	16 Q9JZ16	Q9JZ16 methylobact
36	169.5	13.4	193	16 Q9ZAZ5	Q9ZAZ5 neisseria m
37	169	13.3	213	16 Q9J081	Q9J081 listeria in
38	168.5	13.3	201	16 Q98E51	Q98E51 rhizobium l
39	168	13.3	200	16 Q9KP80	Q9KP80 vibrio chol
40	167.5	13.2	216	16 Q8UBR0	Q8UBR0 agrobacteri
41	163	12.9	201	16 Q9C3N8	Q9C3N8 pasteurella
42	161.5	12.8	177	16 Q9RTY5	Q9RTY5 delinococcus
43	159.5	12.6	182	2 Q938C8	Q938C8 mycobacteri
44	159	12.6	190	16 Q995J2	Q995J2 staphylococ
45	158.5	12.5	211	16 Q8YCW7	Q8YCW7 brucella me

ALIGNMENTS

RESULT 1	ID	Q949D1	PRELIMINARY;	PRT;	257 AA.
AC	Q949D1				
DT	01-DEC-2001	(TREMBLrel, 19, Created)			
DT	01-DEC-2001	(TREMBLrel, 19, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel, 21, Last annotation update)			
DE	Hypothetical 26.4 kDa protein.				
GN	C740ERIPDM.				
OS	Oryza sativa (Rice).				
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
CC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
CC	Ehharitoidae; Oryzae; Oryza.				
OX	NCBI_TaxID=4530;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21329048; PubMed=11435398;				
RA	Mayer K., Murphy G., Tarchini R., Wambutt R., Volckaert G., Pohl T.,				
RA	Duesterhoeft A., Stiekema W., Entian K.D., Terry N., Lemke K.,				
RA	Haese D., Hall C.R., van Dodeveld A.M., Tingey S.V., Mewes H.W.,				
RA	Bayan M., Bancroft I.;				
RT	"Conservation of microstructure between a sequenced region of the				
RT	genome of rice and multiple segments of the genome of Arabidopsis				
RT	thaliana."				
RL	Genome Res. 11:1167-1174(2001).				
DR	EMBL; AJ307662; CAC39061.1; -				
DR	InterPro; IPR000573; Aconitase_C.				
DR	Pfam; PF00694; Aconitase_C; 1.				
KW	Hypothetical protein.				
SO	SEQUENCE 257 AA; 26443 MW; 7EC420ECF735FB66 CRC64;				

Query Match 78.2%; Score 990.5; DB 10; Length 257;

Best Local Similarity 76.7%; Pred. No. 4.6e-76;

Matches 197; Conservative 18; Mismatches 33; Indels 9; Gaps 2;

QY	1	MAAALSTAVSTA-----ALLAPRAPTSAFLRBSOLTRHLSKC---RAGSIVPA 51
DB	1	MAAALSTAVSTAALLAPRAPTSPSTFRFRSMVAALCRPALCKHHSRPLTAVAAA 60

```

QY 52 AAAAAAGSSPSSAVFHGECFVGVNDITDQIIPAEHLTLVPSKPDPEYRKLGSFAFAGLP 111
    ||||| | : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 AAAAAAGDSTAGVFHGEFVGVNDITDQIIPAEHLTLVPSKPDPEYRKLGSFAFAGLP 120
    ||||| | : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 112 SAAPFPFVAVGESSRYAIVGANGCGSSREHAPVALGAAGARVAEGYARIFERN 171
    : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TAAFPFPFVAVGEBTTRVAVIGANGCGSSREHAPVALGAAGARVAEGYARIFERN 180
    : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 172 SVATGEVYPLELDVGVAMKECKTGDDVTVVLANSVFNHSGEYKLPFGDAGPVIENG 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 SVATGEVYPLELDVGVAMKECKTGDDVTVVLANSVFNHSGEYKLPFGDAGPVIENG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 232 GIFAYARKTGMIASKAA 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GIFAYARKTGMIASKAA 257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
Q92W85 PRELIMINARY: PRT; 251 AA.
092W85
AC Q92W85;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 3-isopropylmalate dehydratase, small subunit (Putative
DE 3-isopropylmalate dehydratase small subunit).
GN At2G43090 OR F1AB2.3/At2G43090.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana."
RT Nature 402:761-768(1999).
RL
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RX Lin X.;
RA submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koeseema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "Full Length cDNA of gene F1AB2.3/At2G43090 (GI:3763918).";
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RA "Full Length cDNA of gene F1AB2.3/At2G43090 (GI:3763918).";
RT

```

```

RL submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanaken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC004450; AAC64298.1; -
DR EMBL; AY035158; AAK59662.1; -
DR EMBL; AY063029; AAL34203.1; -
DR EMBL; AC006224; AAM15163.1; -
DR InterPro; IPR000573; Aconitase_C.
DR Pfam; PF00694; Aconitase_C.1.
SQ SEQUENCE 251 AA; 26790 MW; 62EBB1B6059ED272 CRC64;

Query Match 56.1%; Score 710.5; DB 10; Length 251;
Best Local Similarity 59.5%; Pred. No. 2,3e-52;
Matches 153; Conservative 27; Mismatches 62; Indels 15; Gaps 6;

QY 1 MAALSGTAVSTAAIAPRAPTSAFIRRSQITCRHLSLTCORAGSTVPAAAAAAGSS 60
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MAASIQSANPTLSKRLASPNKPSFATFRSPL--RNMSTSY--ASNFKPLVSRASSSF 56
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 SPSSA-----VFHGEFVGVNDITDQIIPAEHLTLVPSKPDPEYRKLGSFAFAGLP 113
    || : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57 VTRSAAEPEKRTFHGLCTVGVNDITDQIIPAEHLTLVPSKPDPEYRKLGSFAFAGLP-A 115
    || : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 AVTPFPFVAVGESSRYAIVGANGCGSSREHAPVALGAAGARVAEGYARIFERN 173
    : || || : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 SYKERFVQPGEMKTKYXIIIGENGCGSSREHAPVALGAAGARVAEGYARIFERN 175
    : || || : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 ATGEVYPLELDVGVAMKECKTGDDVTVVL--ANSVFHNHSGEYKLPFGDAGPVIENG 231
    ||||| ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 ATGEVYPLD--SEYRVDCDETTDVAIVELRESGDIILNHHTKEKRLPIGDAGVIDAG 234
    ||||| ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 232 GIFAYARKTGMIASKAA 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 235 GIFAYARKTGMIASKAA 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
Q9LYT7 PRELIMINARY: PRT; 253 AA.
09LYT7
AC Q9LYT7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 3-isopropylmalate dehydratase-like protein (Small subunit) (Putative
DE 3-isopropylmalate dehydratase small subunit).
GN F17J16_40 OR AT3G58990.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,

```


RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the *Pyrococcus furiosus* genome."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE010207; AAL81063.1; -
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 164 AA; 18257 MW; F959DB935A677034 CRC64;

Query Match 21.0%; Score 265.5; DB 17; Length 164;
 Best Local Similarity 39.3%; Pred. No. 6,7e-15;
 Matches 68; Conservative 27; Mismatches 59; Indels 19; Gaps 8;

OY 69 GECFVGDNDITDQIIIPAHLTLVPSKPDYERKLSFAFAGL-PSAAYPPFPVAPGESS 127
 DB 5 GRAMKYGNDITDVIIPARLYN--TSDPKR--LAQHVLDDLPFRSK--MKPGD--- 53
 OY 128 RYATVGGANFGCCSSREHAPVALGAGARAIYAGYARIFRNSVATGEVPLELTDVGA 187
 DB 54 ---ITVAGENFGCCSSREHAPLAKAGVSAVIAKSFARIFYRNAINIG--LPI-LEAPE 107
 OY 188 AMKECKTGDVTVDLANSVFINTSGKEYKLPIGD-AGPYIAGGIFAYARK 239
 DB 108 AVERIETGDEIDEIFSTGEIRNLTKGEVYHANPPEEFIMELIKAGLVEWAKR 160

RESULT 6
 OY7EE1 PRELIMINARY; PRT; 163 AA.

AC 097EE1;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 3-Isopropylmalate dehydratase, small subunit.
 GN CAC3172.
 OS *Clostridium acetobutylicum*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hilti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium *Clostridium acetobutylicum*."
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL: AE007812; AAK81109.1; -
 DR InterPro: IPR000573; Aconitase_C.
 DR Pfam: PF00694; Aconitase_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 163 AA; 18025 MW; 1A4D16171A32AAE2 CRC64;

Query Match 20.5%; Score 260; DB 16; Length 163;
 Best Local Similarity 38.3%; Pred. No. 1.9e-14;
 Matches 64; Conservative 22; Mismatches 43; Indels 38; Gaps 6;

OY 68 HGECEVVGNDITDQIIIPAHLTLVPSK-----PDYERKLSGFARAGLPSAAYP 116
 DB 4 NQDVLKYGNDITDVIIPARLYNTSVPEELAKHCKMEDLDVFLKLTKT----- 51
 OY 117 TFEVAPGESSRYATVGGANFGCCSSREHAPVALGAGARAIYAGYARIFRNSVATG 176
 DB 52 -----GD-----ITVGGRNFGCCSSREHAPICIKAGVSCVIAKSFARIFYRNSINIG 99
 OY 177 EYVPLELTDVGAAMKECKTGDVTVDLANSVFINTSGKEYKLPIGD 223
 DB 100 --FPI-LECCEAVNDASTGDKLEVDIFEGILNVTILNKEYKQPPD 143

RESULT 7

OY90Z06 PRELIMINARY; PRT; 164 AA.

AC OY90Z06;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE 3-Isopropylmalate dehydratase, small subunit (LEUD-1).
 GN LEUD-1 OR PAB0892.
 OS *Pyrococcus abyssi*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORSAY;
 RA Helling R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
 RT structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ248287; CAB50256.1; -
 DR InterPro: IPR000573; Aconitase_C.
 DR Pfam: PF00694; Aconitase_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 164 AA; 18217 MW; EBD01AE526C02D5D CRC64;

Query Match 20.4%; Score 258.5; DB 17; Length 164;
 Best Local Similarity 39.0%; Pred. No. 2.6e-14;
 Matches 67; Conservative 26; Mismatches 62; Indels 17; Gaps 7;

OY 69 GECFVGDNDITDQIIIPAHLTLVPSKPDYERKLSFAFAGLPSAAYPPFPVAPGESSR 128.
 DB 5 GRAMKYGNDITDVIIPARLYN--TSDPKR--LAKHVLDDLPFRSK--MKPGD--- 53
 OY 129 YATVGGANFGCCSSREHAPVALGAGARAIYAGYARIFRNSVATGEVPLELTDVGA 188
 DB 54 ---ITVAGENFGCCSSREHAPLAKAGVSAVIAKSFARIFYRNAINIG--LPI-LEAQA 108
 OY 189 WKCECKTGDVTVDLANSVFINTSGKEYKLPIGD-AGPYIAGGIFAYARK 239
 DB 109 VDRIFETGDELEVDFSSGEIRNLTKGEVYHANPPEEFIMELIKAGLVEWAKR 160

RESULT 8
 OY8TX94 PRELIMINARY; PRT; 170 AA.

AC OY8TX94;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 3-Isopropylmalate dehydratase small subunit.
 GN LEUD_1 OR MK0781.
 OS *Methanopyrus kandleri*.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyrtaeae;
 OC Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Stesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Matala D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
 RA Malynch A.G., Koonin E.V., Kozayavkin S.A.;
 RT "The complete genome of hyperthermophilic *Methanopyrus kandleri* AV19
 RT and monophyly of archaeal methanogens."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL: AE010370; AAM01995.1; -
 KW Complete proteome.
 SQ SEQUENCE 170 AA; 18686 MW; E329CD3D36367F6 CRC64;

Query Match 20.1%; Score 254.5; DB 17; Length 170;
 Best Local Similarity 37.3%; Pred. No. 6e-14;
 Matches 66; Conservative 26; Mismatches 68; Indels 17; Gaps 7;

	QY	66 VFHECECVVDNIDTFOIIPAEHLTVPSKPDREKLGSFAAGLPSANYPPTFPVAPGE	125
	Db	4 VIRARNAVFDDDIDTDIIIGRGLT--TTQDPPE---LAHWMEG-ADPEFF-----EK	50
	OY	126 SSRATIVGGANFGCGSSSRHAPVALGAAGAARVAIEGYARIIFRNVSATGVPLELTLD	185
	Db	51 VREGDVIAAKNKNGCGSSSRHAPIALKAGACLVVRSEAFIRFYNAIMLG--LELLVCP	108
	OY	186 VGAAKECKTDVVTVDLANSVFINHSGKEYLKPIGD-AQSVTEAGSFEAVARRKTG	241
	Db	109 -GYDDADEFDGQGIENVLRSEYVKNLDTGETELEKPLPDMRMILLEGVLIELIKREG	164
	RESULT 9		
	QB8DK1		
	ID	PRELIMINARY:	PRT; 161 AA.
	OCB8DK1		
	DT	01-JUN-2002 (TREMBLrel. 21, Created)	
	DT	01-JUN-2002 (TREMBLrel. 21, last sequence update)	
	DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)	
	DE	3-Isopropylmalate dehydratase small subunit.	
	GN	LEUD OR TTE0018.	
	OS	Thermoanaerobacter tengcongensis.	
	CC	Bacteria, Firmicutes; Bacillus/Clostridium group; Clostridia;	
	CC	Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter	
	OX	NCHI_TaxID=119072;	
	RN	[1]	
	RP	SEQUENCE FROM N.A.	
	RC	STRAIN-MBAT / JCM11007;	
	RX	MEDLINE=21992816; PubMed=11997336;	
	RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,	
	RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,	
	RA	Tan H., Chen R., Wang J., Yu J., Yang H.;	
	RT	"A complete sequence of T. tengcongensis genome.";	
	RL	Genome Res. 12:689-700(2002).	
	DR	EMBL: AEO12976; AAA23335.1; .-	
	KW	Complete proteome.	
	SQ	SEQUENCE 161 AA; 17888 MW; 4C4A2BCF9CC2B6BF CRC64;	

RX MEDLINE-21929760; PubMed-11932238; Endrizzi M.G., Macdonald P.,
 RA Galscan J.E., Nusbbaum C., Roy A., Endrizzi M.G., Brown A.,
 RA Fitzhugh W., Calvo S., Engels R., Smlnov S., Atmoo D., Brown A.,
 RA Allen N., Naylor J., Strange-thomann N., Deatellano K., Johnson R.,
 RA Linton L., McKean P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kueltner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pitchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of *Methanoscarcina acetivorans* reveals extensive metabolic
 RT and physiological diversity.";
 RT genome Res. 12:532-542(2002).
 RL genome Res. 12:532-542(2002).
 DR EMBL; A6010678; AAM03655.1; -.
 KW Complete proteome.
 QO Stouffence 162 AA; 17825 MW; 2D609D9BA176A228 CFC64;

Query Match 19.9%; Score 252; DB 17; Length 162;
Best local Similarity 35.8%; Pred. No. 9.1e-14;
Matches 62; Conservative 31; Mismatches 60; Gaps 7.

Dy
Qy
Dy
Qy
Db
Qy
Dy
Qy
Dy
Qy
Dy

69 GCEFFVGGNDIDDIITPEHNTLTPSPDEDRKLGSFAPAG-PSAAATPFPVAPGESS 127
::
3 GRAMKGGDDVDIDAIPGRY--LIENPGE--LAAYTEGVGPDOFA-----KKVM 48

128 RVALIYGANFECCSSRRHPALVGAAGAARAIYAGEAATPIFRNSAVNGVEYPLELDVG 187
::
49 ENDIVVAGNSECCSSRRHPALVLAGSKSVCAIKSFALIFRMAINIG-VPLYECNPT- 106
::
188 AMKECTGDVTDLANSVFINTSHGEYELAKTIGD-AQPYLEAGGIAPARK 239
::
107 --DRIDSGELEVEDSTSDIONITRGTYOATPLPFREAVLEDEGLLEVRK 157

RESULT 11			
08ZM36			
ID	08ZM36	PRELIMINARY:	PRT; 161 AA.
AC	08ZM36;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	3-isopropylmalate dehydratase small subunit (leud).		
GN	PAE191.		
OS	Pyrobaculum aerophilum.		
OC	Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;		
OC	Thermoproteaceae; Pyrobaculum.		
OX	NCBI_TaxID=13773;		
RN	[1]		
RF	SEQUENCE FROM N.A.		
RP	STRAIN-IM2 / ATCC 51768 / DSM 7523;		
RC	Pubmed=11792869;		
RA	Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,		
RA	Miller J.H.;		
RT	"genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum		
RT	aerophilum.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).		
DR	EMBL; AE009851; AAL3866.1; -.		
DR	InterPro: IPR000573; Aconitase_C.		
DR	Pfam: PF00694; Aconitase_C.1.		
DR	Complete proteome.		
QC	SEQUENCE 161 AA; 17175 MW; 440B9AA6AF3F45A2 CRC64;		

Query Match	18.28;	Score 231;	DB 17;	Length 161;
Best Local Similarity	36.48;	Pred. No. 5.5e-12;		
Matches 63; Conservative	29;	Mismatches 63;	Indels 18;	Gaps 6;

```

QY      69 GEEFVWGNDITDQITPAHEHLLTPVSPKDEYKRLGSEFARGLPSAITYTPPVAPGESSR 128
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5 GRALVYGDXIDTIDYITPAKLYVT-----DPALGGHAMEPL-DEEP-----KKAK 50
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY     129 YALITGGANFGCGSSREHAPVALCAGARALVAEGARITFFENSVATGEVYPLELTDVGA 188

```

```

Db      51 GAVLVAGRAFGMGSSREQALALKGAGVLAAVAESFARIFFFRNAIIVG--LPV-LQAPGI 107
QY      189 WKCKTGQVYVDLANSVFINTSGEYKLPKI-GDAGVIAGGIFFAVART 240
      108 REKVKGDEVELDEGVGIVNITTGVEIVGKPLRGLPEILKAGGLNLYKNS 160

RESULT 12
O9RT10  PRELIMINARY:      PRT:      208 AA.
ID      O9RT10
AC      O9RT10;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      3-isopropylmalate dehydratase, small subunit.
GN      DR1784.
OS      Deinococcus radiodurans.
OC      Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OX      Deinococcaceae; Deinococcus.
NCBI_TaxID=1299;

RN      SEQUENCE FROM N.A.
[1]
RC      STRAIN-R1;
MDLINE=20036896; PubMed=10567266;
RA      White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA      Dodson R.J., Qian D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA      Moffat K.S., Haft D., Jiang L., Pamphile W., Crosby M., Shen M.,
RA      Vamatheva K.S., Lam P., McDonald L., Uterback T., Zalewski C.,
RA      Makarov K.J., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA      Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA      Fraser C.M.;
RT      "Genome sequence of the radioresistant bacterium Deinococcus
RT      radiodurans R1."
RL      Science 286:1571-1577(1999).
DR      EMBL; AE002019; AAF11332.1; -.
DR      TIGR; DR1784; -.
DR      InterPro: IPR000573; Aconitase-C.
KW      pfam; PF00694; Aconitase_C; 1.
KM      Complete proteome.
SQ      SEQUENCE 208 AA; 22652 MW; F26C4BB27FEBD1E7 CRC64;

Query Match      18.1%; Score 229.5; DB 16; Length 208;
Best Local Similarity 38.3%; Pred. No. 1e-11;
Matches 64; Conservative 21; Mismatches 59; Indels 23; Gaps 6;

QY      76 DNIDTDQIIIPAHLTVPSKPDYRKLGSAFAAGLPASAYPTPP---VAGGESSRAIT 132
      10 DHINTDEIIPARHLT-----TDVSEELAKYAMED-----YDKDFVRRVQPD-----II 53
Db      133 VGSANFGGSSREHAPVALGAAGARAIYEGYARIFFRNSVATGEVYPLELTDVGAKREC 192
      54 VAGADFGGSSREHAWALRGASAYAIAPFAIIRYNSINNGF---LALECGITIELF 110
QY      193 KTGADVVDLANSVFINTSGEYKLPKIGD-AGPVIENGAGIFAYAR 238
      111 QDGEAEALDLKGTIRNRTKEISFVVPQFALDVGKAGGMLYMK 157
Db

RESULT 13
O9RTF7  PRELIMINARY:      PRT:      164 AA.
ID      O9RTF7
AC      O9RTF7;
DT      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      3-isopropylmalate dehydratase.
GN      LEUC OR MA1223.
OS      Methanosarcina acetivorans.
OC      Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC      Methanosarcinaceae; Methanosarcina.
OX      NCBI_TaxID=2214;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN-CZA / ATCC 35395 / DSM 2834;
RA      MEDLINE=21929760; PubMed=11932238;
RA      Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA      FitzHugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA      Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA      Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA      Zimmer A., Barber R.D., Cann I., Graham D.E., Guss A.M.,
RA      Hedderich R., Ingram-Smith C., Kuettnet H.C., Krzycki J.A.,
RA      Leigh J.A., Li W., Liu J., Mukhopadhyay B., Keefe J.N., Smith K.,
RA      Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA      Ferry J.G., Jarrell K.F., Jing H., Macario A.J.D., Paulsen I.,
RA      Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA      Metcalf W.W., Birren B.;
RT      "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT      and physiological diversity."
RL      Genome Res. 12:532-542(2002).
DR      EMBL; AE010790; AAM04642.1; -.
DR      Complete proteome.
KM      Complete proteome.
SQ      SEQUENCE 164 AA; 17963 MW; 7D4660C68589AC0 CRC64;

Query Match      18.0%; Score 227.5; DB 17; Length 164;
Best Local Similarity 37.0%; Pred. No. 1.1e-11;
Matches 64; Conservative 25; Mismatches 61; Indels 23; Gaps 7;

QY      69 GECFVGNIDPDQIIIPAHLTVPSKPDYRKLGSAFAAGLPASAYPTPFVAPGESSR 128
      7 GAWMKFGDINDDAIIPKYL-----RTRDMQIFGTNAMEGIDPEF--TKRAKPGD---- 55
Db      129 VAIIVGANGGCGSSREHAPVALGAAGARAIYEGYARIFFRNSVATGEVYPLELTDVGA 188
      56 --IIVAGTNGGCGSSREHAPVALKSGIACIASFAIFFRNAINIG--LPLMDADV-- 109
QY      189 WKCKTGQVYVDLANSVFINTSG--KEYKLPKIGDAGPYIENGAGIFAYARK 239
      110 --ECQGDRIKYLKRGVIVPEKGIFFGNKLPDP--LLDILNDGGLVAHHKK 158
Db

RESULT 14
O9WYC8  PRELIMINARY:      PRT:      166 AA.
ID      O9WYC8
AC      O9WYC8;
DT      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      3-isopropylmalate dehydratase, small subunit, putative.
GN      TM0292.
OS      Thermotoga maritima.
OC      Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX      NCBI_TaxID=2336;
RN      [1]
RC      SEQUENCE FROM N.A.
RC      STRAIN-MSB8 / DSM 3109;
MDLINE=99287316; PubMed=10360571;
RA      Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA      Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA      McDonald L., Uterback T.R., Malek U.A., Linher K.D., Garrett M.M.,
RA      Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA      Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA      Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT      "Evidence for lateral gene transfer between Archaea and Bacteria from
RT      genome sequence of Thermotoga maritima."
RL      Nature 399:323-329(1999).
DR      EMBL; AE001711; MAD35380.1; -.
DR      TIGR; TM0292; -.
DR      InterPro: IPR000573; Aconitase-C.
DR      Pfam: PF00694; Aconitase_C; 1.
KM      Complete proteome.
SQ      SEQUENCE 166 AA; 18734 MW; D8ACF34020511A4 CRC64;

Query Match      18.0%; Score 227.5; DB 16; Length 166;
Best Local Similarity 34.2%; Pred. No. 1.1e-11;
Matches 64; Conservative 26; Mismatches 60; Indels 37; Gaps 7;

```


Search completed: March 31, 2003, 09:08:53
Job time : 37 secs

```

QY 66 VFHGEFVVDNIDTDQIIPAEHLTLVPSKPDPEYRK-----LGFAPAGLPSAAYPTPF 119
DB 1 MINGRAWKFGDNISTDHIAPGRYFHL-RNNLELAKHVLEDMEDFA----- 46
QY 120 VAFGESSRAYIIVGANGFCSSREHAPVALGAAGARAIYAEGARIFFRNSVATGEVY 179
DB 47 ----KKYQKGDILVAKNGGICSSREHARIITKINGVSCIYAKSFARIFFRNAIVNG-LP 101
QY 180 PLELTDVGAMKECKTGDVYTVDLANSVFTNHTSGKEYKLKPIGDAGP-----VIEAGIIF 234
DB 102 VIELKEV--DEINOGDELEIDLKLVKLTGKEYRFTPI---PFELLEILKEDGIV 154
QY 235 AYARKTG 241
DB 155 NYLKKHG 161

```

RESULT 15

```

Q8TJM9 PRELIMINARY; PRT; 166 AA.
ID 08TJM9
AC 08TJM9;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 3-Isopropylmalate dehydratase.
GN LEUD OR MA3751.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OX Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE01086; AAM07104.1; -.
KW Complete proteome.
SQ SEQUENCE 166 AA; 18151 MW; 4BC4ED4738FDE014 CRC64;

```

Query Match 17.8%; Score 225.5; DB 17; Length 166;

Best Local Similarity 33.7%; Pred. No. 17e-11; Mismatches 49; Indels 45; Gaps 8;

```

QY 69 GECFVVDNIDTDQIIPAEHLTLVPSKPDPEYRKLSGFAFAGLPSAAYPTPFVAPGESSR 128
DB 8 GRVWKFGDDIDTDVITIPGKYL-----RTKDWQIFAHAMEGI-APEFTKKAKPGD----- 56
QY 129 YAIIVGANGFCSSREHAPVALGAAGARAIYAEGARIFFRNSVATGEVYPLELTDVGA 188
DB 57 --IIVAGENFGCSSRQALAKHAGIACVYAKSFARIFFRNAIVNG--LEPLMEADI-- 110
QY 189 WKECKTGDVYTVDL-----ANSVFINHTSGKEYKLKPIGDAGP-----VIEAGIIFA 235
DB 111 --EQQEGDETEVDLKGAVVPGKGVF-----VGNKLPDFLDDILTIDGGLYS 155
QY 236 YARK 239
DB 156 HRKK 159

```

